IDENTIFICATION OF POTENTIAL EPITOPES FOR THE DESIGN OF HIV-1 VACCINE IN KENYA

DOMTILA WANGUI KIMANI (B.ED.SC)

156/5729/2003

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE (IMMUNOLOGY) IN THE SCHOOL OF PURE AND APPLIED SCIENCES OF KENYATTA UNIVERSITY

NOVEMBER 2008

DECLARATION

I Domtila Wangui Kimani declare that this is my original work and has not been presented for a degree in any other University or any other award.

Signature……………………………Date 1/12/2008……………………

DOMTILA WANGUI KIMANI

This thesis has been submitted for examination with our approval as University supervisors

DR. MICHAEL GICHERU Signed……………………………Date 2-12-08
Department of Zoological Sciences
Kenyatta University

PROF. ZIPPORAH NG'ANG'A Signed……………………………Date 2/12/08
Department of Medical Laboratory Sciences
Jomo Kenyatta University

DR. SAMOEL ASHIMOSI KHAMADI Signed……………………………Date 2/12/08
Center for Virus Research (CVR)
Kenya Medical Research Institute, Nairobi, Kenya

DEDICATION

I dedicate this thesis to my parents James Kimani and Linah Kimani whose discipline and love gave me eyes to see grace.

To my brothers Mike and Steve and my sisters Margy and Brandy for their encouragement, tolerance and support.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my supervisors Dr Michael Gicheru, Dr Samoel Khamadi and Prof. Zipporah Ng'ang'a for their invaluable advice, constant guidance and encouragement throughout the course of the study and thesis writing.

I must specifically acknowledge Dr Samoel Khamadi and Dr Matilu Mwau for their assistance with numerous key ideas especially in epitope prediction and vaccine design.

I am also indebted to Joyceline Kinyua, Nancy Lagat, Mwangi Joseph, Lihana Raphael, Vincent Okoth, Muriuki Joseph and Alex Maiyo of the Kemri HIV lab for their assistance throughout the period and especially for the technical assistance in the laboratory.

Finally, many thanks to Kenyatta University and in particular Department of Zoological Sciences for the opportunity to undertake postgraduate study and to KEMRI for the opportunity to be part of them in research and for the financial sponsorship that made this study a success.

This work would not have been possible without the help of you all. May God bless you abundantly.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

ABI	Applied BioSystems International
AIDS	Acquired Immunodeficiency Syndrome
ARV	Antiretroviral
BLAST	Basic Local Alignment Tool
CCD	Charged Coupled Device
CCR5	CC Chemokine Receptor-5
CDC	Center for Disease Control and Prevention
CD4	Cluster Differentiation 4
CD4+ T cell	T lymphocyte positive for lymphocyte marker
CD8	Cluster Differentiation 8
CD8+Tcell	Cytotoxic T lymphocytes
CDNA	Complementary Deoxyribonucleic Acid
CDR 1-3	Complementary Determining Regions 1-3
CR 1, 2, 3	Complement Receptor 1, 2, 3
CMI	Cell Mediated Immunity
CRF	Circulating Recombinant Forms
CVR	Center for Virus Research
CXCR4	CXC chemokine receptor-4
DNA	Deoxyribonucleic Acid
DCs	Dendritic Cells
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme Linked Immunosorbent Assay

Env	Envelope
FDC	Follicular Dendritic Cells
Fasl	Fas ligand
Gag	Group Associated Genes
HIV-1, -2	Human Immunodeficiency Virus types 1,-2
HLA	Human Leucocyte Antigen
IFA	Freud's Incomplete Adjuvant
Ig	Immunoglobulin
IL	Interleukin
Inr	Initiator
ISCOMS	Immunostimulatory Complexes
KEMRI	Kenya Medical Research Institute
LTR	Long Terminal Repeat
MADH	Mandera District Hospital
MHC	Major Histocompatibility Complex
MYDH	Moyale District Hospital
Nef	Negative Factor
NSI	Non-syncytium Inducing
PCR	Polymerase Chain Reaction
Pol	Polymerases
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RNAPII	Ribonucleic Acid polymerase II

RPM	Revolutions per Minute
SI	Syncytium Inducing
SIV	Simian Immunodeficiency Virus
SSA	Sub-Saharan Africa
Tat	Transcriptional transactivator
TAE	Tris Acetic Acid EDTA
Taq	Thermus aquaticus
TBE	Tris Borate EDTA
T-H1 and II	T-Helper 1 and II
TKMH	Turkana Mission Hospital
TLHC	Turkana Lobiding Health Centre
TSR	Template Suppression Reagent
TNF α and β	Tumour Necrosis Factor α and β
µL	Microlitre
UNAIDS	Joint United Nations Programme on HIV/AIDS
URFs	Unique Recombinant Forms
Vif	Viral infectivity factor
Vpu	Viral protein, Unknown
WHO	World Health Organization

ABSTRACT

Human Immunodeficiency Virus is a major public health problem, socio-economic burden and a serious threat to development globally. HIV exhibits an extremely high capacity for genetic variation with rapid turnover of virions. The heterogeneity of HIV may ultimately lead to increased viral fitness in the face of pharmacological, immunological or other environmental selection pressures. The high genetic diversity of HIV-1 continues to complicate measures for the design of an effective vaccine. An effective HIV-1 vaccine would have to stimulate a range of host defenses, including mucosal, innate immunity, neutralizing antibodies and cell-mediated immunity. The variability of the HIV-1 envelope region and the inaccessibility of potentially neutralizing epitopes on primary isolates continue to hamper the development of vaccines for HIV. The focus has shifted to the induction of CD8+ cytotoxic T-lymphocytes (CTLs), which have been shown to play an important role in the control of HIV infection. The ability of a vaccine to induce responses directed at a particular group of epitopes is of interest because it is easier to assess the possible efficacy of any resultant immune response. The knowledge of epitopes is critical in the precise evaluation of the strength and quality of CTL responses that could be induced by vaccine candidates and it would be helpful in identifying immunologically silent regions of a vaccine so that they can be omitted from future constructs. The use of epitopes that are conserved across clades could improve the breadth of induced responses. In this study, the aim was to identify the HIV-1 subtypes circulating in Northern Kenya and to identify conserved immunogenic epitopes that can be used to design a multiepitope cross-clade candidate vaccine to be used in Kenya. Phylogenetic analysis of the generated *gp41* sequences showed that 44% of the sequences generated from the three districts were HIV-1 subtype A1, 45% were HIV-1 subtype C and 11% were HIV-1 subtype D. Samples from Moyale indicated 36% of subtype A1, 55% were HIV-1 subtype C and 9% were HIV-1 subtype D while from Mandera 67% were HIV-1 subtype A1, 33% were HIV-1 subtype C. In Turkana the most dominant HIV-1 subtype was A1 (58%), HIV-1 subtype C was 25% and HIV-1 subtype D was 17%. There was a significant difference in the pattern of subtypes circulating in the three regions in that both Turkana and Mandera had subtype A1 as the predominant subtype while subtype C was the dominant subtype in Moyale. For epitope determination, sequences generated from the *env gp41*, *env gp120 (C2V3)* and the *p24 gag* regions of the HIV-1 genome were analysed. The generated sequences were translated to amino acid sequences using the Translation for publication software and aligned using ClustalW version 1.81 software to determine the areas of the sequences that were conserved and therefore relevant for design of a candidate vaccine. The identified conserved epitopes were further analyzed using the SYFPEITHI bioinformatics tool to identify class 1 restricted T-cell epitopes and their immunogenicity. A total of 80 epitopes from *gp* 120, 41 from *gag* and 37 from *gp* 41 were identified. The identified epitopes were used to construct a super-epitope that can be used to design a HIV-1 candidate vaccine. Information generated from this study can be used to address the challenges of HIV-1 viral diversity in the development of vaccine candidates. This can be part of the long term effort to build a panel of subunit vaccines that can be used in design of an efficacious vaccine.

CHAPTER ONE: INTRODUCTION

1.1 Background Information

Acquired Immune Deficiency Syndrome (AIDS), first documented in 1981 (Gallo *et al.*, 1983; Kandathil *et al* 2005) and was subsequently attributed to infection with Human Immunodeficiency Virus (HIV). HIV is the etiological agent of the slow progressing Acquired Immunodeficiency Syndrome (AIDS). AIDS results from depletion of CD4 T cells, the major co-ordinating cells of the immune system. Entry of HIV-1 into target cells requires the binding of the external envelope glycoprotein *gp120* to both CD4 molecule and one of the several chemokine receptors, recently discovered to function as co-receptors (Cocchi *et al.*, 1995; Zhang *et al.*, 2002; Feng *et al.*, 2003;).

The Human Immunodeficiency Virus (HIV) has brought about a global epidemic far more extensive than was predicted a decade ago. Though the rates of infection is declining for the majority of the developing countries, Sub-Saharan Africa remains the hardest-hit (Buve *et al* 1995). WHO estimates 15,000 people become infected every day despite progress in prevention and powerful drug combinations to treat HIV infections (http://www.unaids.org/Epidemic_update/).

Over 90% of new infections occur in developing countries for which the recent medical advances are not immediately applicable or affordable (www.iavireport.org, 2005). In most countries the age of greatest morbidity and mortality from AIDS is the most economically productive bracket (20-39 years: www.iavireport.org, 2005). Nations have to grapple with the unprecedented depletion of the work force.

Highly Active Antiretroviral Therapy (HAART) has prolonged survival, reduced the level of viral load and alleviated suffering. However, a large majority of those infected with HIV in most developing countries cannot afford antiretrovirals and treatment is life long. In addition the drugs available have serious side effects, compliance and drug interactions are a serious concern in most recipients and the clinical benefits are usually short lived due to the emergence of resistant strains (Kalams *et al*., 1999). Besides, proper administration of HIV medication, overall HIV management requires major investment in health care infrastructure, skilled health care and the of state art laboratory for diagnosis, viral load determination, drug resistance and blood level monitoring. Therefore, the urgency of the need for HIV vaccines cannot be over emphasised. The only hope for these countries is the development of safe, effective, accessible and preventable prophylactic and/or therapeutic HIV vaccine (McMichael, 2006).

Vaccines are easy to use and relatively cheap (McMichael and Rowland-Jones 2001), their administration is infrequent and can reduce the incidence of disease markedly. Data from non-human primates and T- cell based immunological mechanism of HIV resistant uninfected humans strongly support the belief that a successful HIV vaccine is possible (Hanke and McMichael, 2000; Amara *et al.,* 2001; Shiver *et al*., 2002). To achieve protection, a prophylactic vaccine has to induce both humoral and cell mediated responses (CMI). There has been a considerable effort to develop envelope-based vaccines inducing neutralising antibodies (nAB). Although some new promising approaches to induction of nAB are being developed (Wyatt *et al.,* 1998; Kolchinsky *et* *al.,* 1999), the emphasis of many vaccine designers has shifted to the induction of CMI responses (McMichael and Rowland-Jones, 2001).

The significance of the extensive genetic diversity of HIV isolates and its implication for vaccine design has long been debated. Currently, candidate vaccines are derived from isolates, with the hope that they will be sufficiently cross-reactive to protect against circulating viruses (Zolla-Pazner, 2004). Despite the high variability in HIV, there are currently two approaches to selecting vaccine strains that attempt to contend with the high levels of HIV sequence variation namely use of conserved sequencea and use of contemporary circulating viruses (Gaschen *et al.*, 2002).

An effective way to minimise the degree of sequence dissimilarity between a vaccine and contemporary circulating viruses is to create artificial sequences that are 'conserved' to the viruses. To design such a sequence involves use of a consensus sequence based on the most common amino acids conserved in each position in an alignment (Novitsky *et al.,* 2002; Korber *et al.,* 2001). Consensus sequences and strains from variable strains could be used in a polyvalent vaccine approach (Korber *et al.,* 2001). Conserved proteins from different subtypes can be more closely related than variable protein from the same subtype and this could be exploited by using a single vaccine strain for conserved proteins and multiple clade-specific strains for variable vaccines specifically designed to target these conserved epitopes. The use of epitopes that are conserved across clades could improve the breadth of induced responses. If successful, this may ultimately be optimised by fine-tuning as subtype-specific vaccines (Gaschen *et al.*, 2002).

The control of viremia and virus escape in HIV infection, have all been associated with cellular immune responses to immunodominant epitopes and mutations or deletions occurring in these epitopes, respectively (Borrow *et al.*, 1997; Goulder *et al.*, 1997; 2001). A comprehensive understanding of HIV epitopes would be a distinct advantage in HIV vaccine development. The ability of a vaccine to induce responses directed at a particular group of epitopes is of interest since it is easier to assess the possible efficacy of any resultant immune response. Epitopes that are relevant for vaccine development are usually identified in infected people (Hanke and McMichael, 2000).

Inclusion of any epitope in the future constructs should ideally be supported by experimental evidence of its ability to induce immune response in vaccinated individuals (Borrow *et al.*, 1997). But the ultimate test for any vaccine would involve the live challenge of vaccinated humans with the wild type HIV (Hanke and McMichael, 2000) though development has been hampered by the high variability of the HIV virus. Given this limitation, vaccination remains the ultimate tool that will complement the preventive programes.

1.2 Problem Statement and Justification

Globally, HIV-1 is extraordinarily variable. The diverging trend of HIV-1 subtypes, existence of mosaic viruses and potential recombinants poses a major challenge in the design and testing of HIV vaccines. The number of unique HIV-1 sequences in public databases has been steadily increasing every year (Altschul *et al.,* 1997). Currently, candidate vaccines are derived from isolates with the hope that they will provide cross protection against circulating viruses. To cope with the diversity, country-specific vaccines are being considered, but evolutionary relationships may be more useful than regional considerations (Goulder *et al*., 2001). Consensus sequences could be used in vaccine design to minimise the genetic difference between vaccine strains and contemporary isolates, effectively reducing the extent of diversity by half. The most effective type of vaccine would include immunogenic regions or epitopes of the HIV-1 genome that are highly conserved across clades and strains of HIV-1 (Rowland -Jones *et al.,* 1998).

This is because use of conserved sequences across clades could improve the breadth of induced responses. Until recently discovery of conserved epitopes in the HIV-1 genome has been hampered by lack of effective tools that would enable researchers or vaccine developers to develop large HIV-1 protein sequence database for vaccine components (Walker *et al*., 1998; Nardin *et al.,* 2001). Although vaccinating individuals against one subtype might confer protection against other subtypes of HIV-1, it is more likely that vaccinating with a single HIV-1 strain may not be a successful means of protecting against challenge by strains belonging to other clades of HIV-1 (De Groot, 2001). Given this degree of diversity, it is widely believed that a vaccine based on a single strains or subtype of HIV-1 will not be successful against the large spectrum of globally circulating HIV-1 variants (Mahmudul and Mahbub, 2003). One solution to the problem of immunizing against many HIV clades may be to search for components of HIV-1 that are highly conserved and to build a vaccine based on those components (Hanke and McMichael, 2000; Davis *et al*., 2000). The bioinformatics methods of sequence analysis are used to reveal the most prospective proteins or protein fragment of an infectious agent as candidates for vaccine design. This approach could help in designing vaccines against diseases where traditional methods are not successful for example when the viral genome reveals the extreme variability and permanent changes of antigenic properties that make it difficult for the selection of molecular targets for medicines and candidate vaccines (Donnes and Elofsson, 2002). Therefore there was need to generate new conserved epitopes that could reduce significantly the number of peptides that have to be synthesised and tested. This study aimed at identifying the conserved epitopes in HIV-1 circulating in Northern Kenya that can be used in vaccine design and development and compare them with already generated epitopes from other regions of the world. Data obtained in this study is important, considering the fact that development of an efficacious vaccine has so far been very elusive and the HIV virus undergoes rapid transformation.

1.3 Research Question

How does the high level of HIV-1 sequence variation in Kenya affect or determine the selection and design of epitopes to be used as HIV-1 vaccine candidates in Kenya?

1.4 Hypothesis

There is no epitope difference in the strains of HIV-1 circulating in Kenya and those identified from other regions of the world.

1.5 Study Objectives

1.5.1 General objective

To identify conserved amino acid epitopes from the immunodominant *env* and *gag* regions of the HIV-1 genome that can be used to develop a cross-clade candidate vaccine in Kenya.

1.5.2 Specific Objectives

i. To determine the subtypes of HIV-1 circulating in Northern Kenya.

ii. To identify conserved amino acid epitopes that are immunogenic and can be used as vaccine targets in different HIV-1 strains circulating in Northern Kenya.

iii. To compare the generated epitopes with the ones already identified in other parts of the world with an aim of determining if these epitopes are relevant for vaccine trials in Kenya.

iv. To design a cross-clade multiepitope candidate vaccine

CHAPTER TWO: LITERATURE REVIEW

2.1 The HIV ~~Virus~~

Human Immunodeficiency Virus (HIV) displays important genetic variability. There are two types of HIV, HIV-1 and HIV-2 (Kandathil *et al.,* 2005). Human Immunodeficiency Virus I (HIV-I) is responsible for the pandemic and was identified in 1983 (Gallo *et al*., 1983; Barre-Sinousi *et al*., 1983). Differences between HIV-1 and HIV-2 are fairly documented in terms of transmissibility, pathogenesis and pattern of spread (De cock *et al*., 1994; Kanki *et al*., 1994). HIV-2 was first detected in West Africa and is significantly prevalent in those regions, Portugal, Korea and Philippines (Damond *et al*., 2001; Machuca *et al*., 1999; Soriano *et al*., 2000).

Human Immunodeficiency Virus is a ribonucleic acid (RNA) virus that belongs to the family of retroviruses, genus Lentiviridae (Jawetz, 2000). The viral genome is carried by two single RNA strands each approximately 10,000 base pairs long (Cheesbrough, 2000). The central nucleocapsid core of the virion contains two copies of the single stranded viral RNA genome, the enzyme reverse transcriptase and intergrase and capsid protein p24 (Kelleher and Bockel, 2005).

Surrounding the core is a protein capsid surrounded by a double-layered phospholipid envelope with embedded surface glycoproteins (*gp*120; Figure 2.1) that protrudes from the surface and *gp*41 that is embedded in the envelope. These enable the virus to attach to and infect the host cells (Cheesbrough, 2000). The genes that encode *gp*120 mutate rapidly producing many antigenic variants.

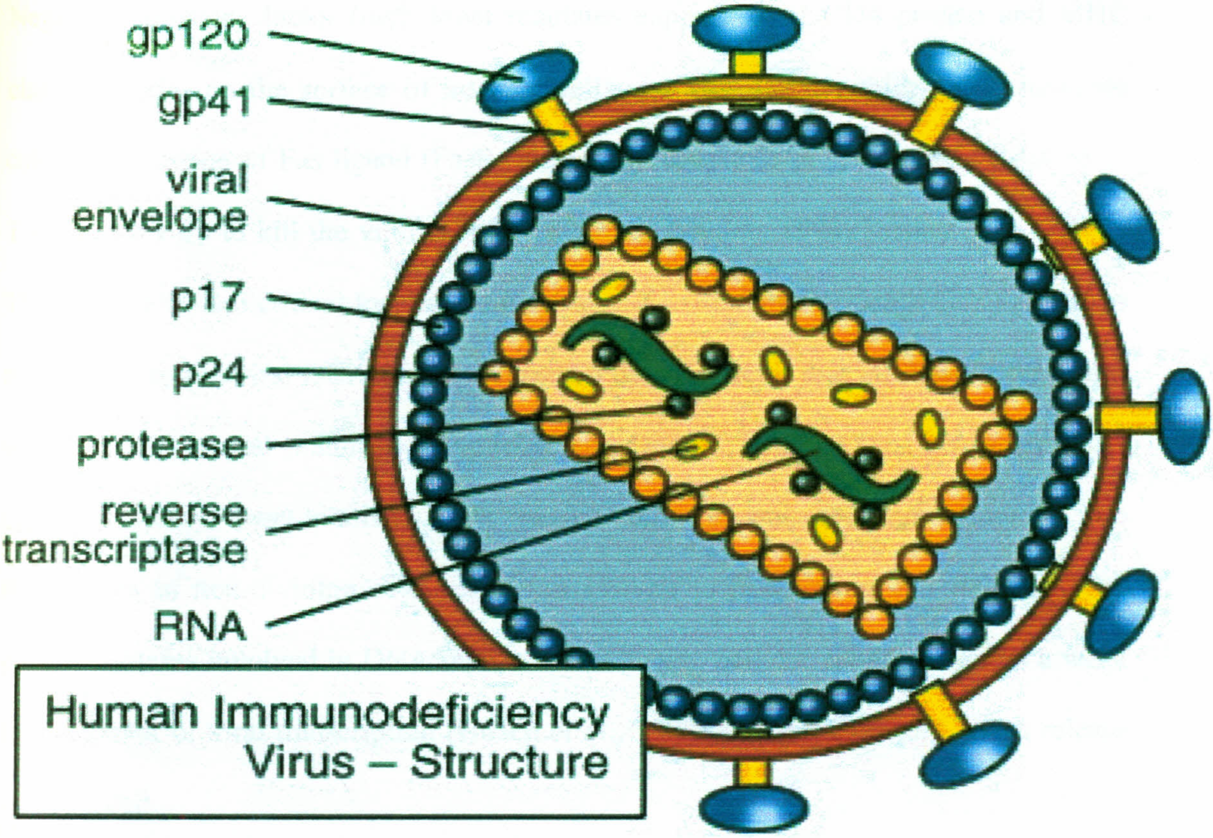


Figure 2.1: The Structure of HIV-1 (http://virology-online.com/viruses/hiv.htm).

2.1.1 The Structure of HIV-1 Genome

The HIV genome consists of two identical molecules of single stranded, positive polarity RNA strand each approximately 10,000 base pairs long (Fred, 1995). Like all retroviruses, the HIV genome (Figure 2.1) consists of structural genes *gag, pol* and *env,* which are in the order of *5'-gag-pol-env-3'* and encode the structural proteins(http://en.wikipedia.org/wiki/Image: HIV_genome.png). In addition, the HIV genome has six regulatory genes namely *tat*, *nef* and *rev* for the replication of the virus and, *vif vpr* and *vpu*, which are accessory proteins not involved in replication (Kelleher and Bockel, 2005; Figure 2.2).

Negative regulatory factor (*nef*) down regulates expression of CD4 protein and MHC class-1 proteins on the surface of infected cells and also has the ability to up regulate cellular expression of Fas ligand (Fasl) resulting in apoptosis of virus-reactive cytotoxic T cell as they try to kill the virally- infected targets (Cullen, 1998; Kestler *et al*., 1991). The *nef* gene is also critical for disease induction. Viral infectivity factor (*Vif*) stabilizes newly synthesized viral DNA and counteracts a cellular factor that inhibits HIV-1 hence is important for viral maturation (Kelleher and Bockel, 2005). Other genes are Viral protein regulatory (*vpr*) involved in the transportation of viral core from cytoplasm into the nucleus in non-dividing cells, promotes cellular differentiation and interacts with cellular proteins involved in DNA repair. It also causes cellular apoptosis and is a weak transactivator of viral transcription (Hirsch *et al*., 1998) and *vpu* enhances virion release from the cell.

The *gag* gene encodes the internal core proteins *p17, p7* and *p24*. The *pol* gene encodes reverse transcriptase, which transcribes RNA genome into DNA (Deacon *et al*., 1995). At both ends of the genome are the long terminal repeats (LTR) that are transcription initiation sites (Triques *et al*., 2000). The 5′ LTR functions like other eukaryotic transcriptional units. It contains downstream and upstream promoter elements, which include the initiator (Inr) and three Sp1 sites. These regions help position the RNA polymerase II (RNAPII) at the site of initiation of transcription and assemble the preinitiation complex (Taube *et al*, 1999).

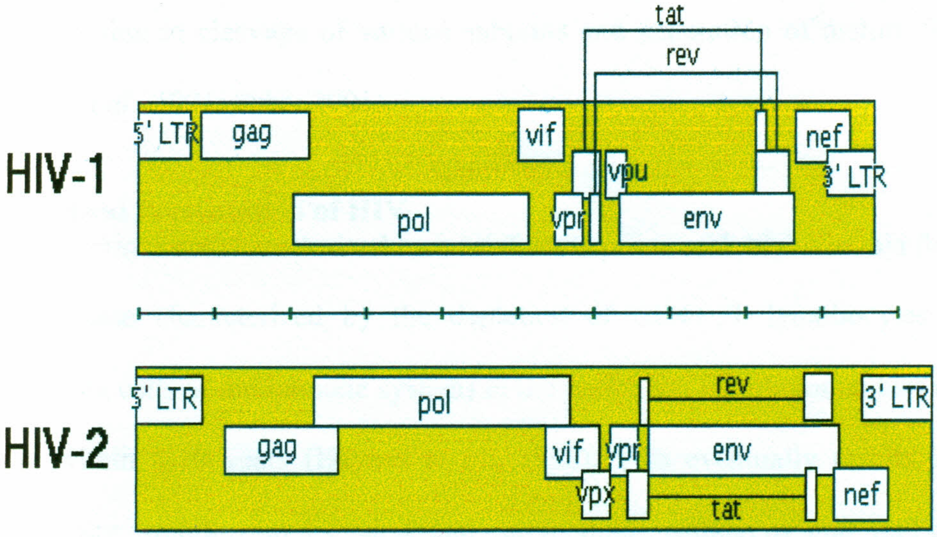


Figure 2.2: Structure of HIV-I and II Genome. (Adapted from the Los Alamos HIV database web.lanl.gov/content/immunology/pdf/2000/intro/GenomeMaps.)

2.1.2 Replication of HIV

HIV replicates within a host cell using RNA as a template to make DNA. To replicate, the virus must transcribe its RNA into DNA using RT (Cheesbrough, 2000). The viral intergrase enzyme is responsible for the intergration of the provirus into the host genome. Viral replication proceeds using host cellular machinery, but is regulated by a number of viral accessory genes such as *nef, vif*, vpr, *tat* and *rev*, some of which are also involved in replication (Kelleher and Bockel, 2005).

The viral messenger Ribonucleic Acid (mRNA) is transcribed from the HIV proviral DNA by the host cell RNA polymerase and translated into several large polyproteins (Mims *et al.,* 1993). Copies of RNA and newly created polyproteins move closer to the cell membrane. New virions begin to form and bulge on the side of the host cell (budding stage). During the budding from the host cell membrane, the viral proteinases become active resulting in cleavage of various subunits and generation of mature form of HIV (Vishnav *et al.,* 1991; Saha, 2001).

2.1.3 Clinical Presentation of HIV

HIV-1 infections are characterized by a brief acute phase with high viremia followed by a chronic phase characterized by the depletion of CD4+ T lymphocytes (the major coordinating cells of the immune system) in the peripheral circulation to less than 20% of normal within 8-10 years (Haynes *et al*., 1996). This eventually results in increased susceptibility to infection by viral and intracellular pathogens and malignancies, in particular non-Hodgkin's lymphoma (Hazenberg *et al* 2000). As immune depletion becomes more profound, infections are dominated by opportunistic pathogens, organisms such as *Pneumocystis carinii*, Cytomegalovirus (CMV), candida and *Mycobacterium avium intracellulae*, which do not cause disseminated progressive infections in immune competent hosts (Kelleher and Bockel, 2005). A healthy, uninfected person usually has 800-1200 CD4+ T cell counts per micro liter (μl) of blood. Infected individuals are vulnerable to opportunistic infections and cancers that typify AIDS, the end stage of HIV disease when CD4+ T cells drop substantially (Hazenberg *et al*., 2000).

2.1.4 Transmission of HIV

HIV is transmitted mainly via mucosal routes (Johnson *et al*., 1999). The different routes include: sexually, by blood and blood products, from mother to child and by other body fluids. Sexual transmission is the most frequent method of transmitting HIV either by heterosexual or by homosexual contact (http://www.cdc.gov/hiv/pubs/facts/transmission.) Inflammatory conditions such as urethritis, epididymitis, or sexually transmitted diseases favour transmission of HIV. This is caused by either increased concentration of HIV in the seminal fluid or by lesions due to infections (Johnson *et al.*, 1999).

Transmission of HIV by blood and blood products has significantly decreased in the developed world since the introduction of HIV screening and recently the introduction of HIV RNA testing of blood and blood products (Cheesbrough, 2000). Sharing needles, syringes or unsafe practices in case of acupuncture, tattooing and piercing can lead to HIV transmission (Cheesbrough, 2000). Human immunodeficiency virus may also be transmitted during pregnancy as early as the first and second trimester or during delivery (De Pasquale, 2003). Factors influencing the transmission of HIV from mother to infant are concomitant STD infection, presence of chorioamniotis at the point of delivery, drug abuse, preterm labour, obstetric procedures and high levels of viremia in the mother (McCutchan, 2000).

Breast-feeding has also been shown to contribute to the transmission of HIV from mother to child and may account for 5-15% of infants becoming infected after delivery (Sagar, 2004). There is no evidence that saliva, sweat or tears that are not contaminated with blood are involved in the transmission of HIV (Cheesbrough, 2000).

2.2 Molecular Heterogeneity of HIV-1

Genetic recombination is part of the normal mechanism of retroviral replication and as such, plays an important role in generation of viral diversity (Najera *et al.,* 2002). The highest genetic diversity of HIV-1 has been found in sub-Saharan Africa where all known HIV-1 subtypes and many of the circulating recombinant forms (CRFs) have been identified, (McCutchan, 2000; Kuiken *et al.,* 2000). The genetic diversity and rapid variation of HIV-1 continues to complicate the development of effective vaccines (Hanke and McMichael, 2000).

Several previously reported studies have shown diverse evolution in different regions of the viral genome (Wong *et al.,* 1997; Zhu *et al.,* 1996) and the reason for this may be attributed to immune system or drug pressure leading to development of immune escape or drug resistant viruses. The envelope gene seems to be subjected to the most extensive genetic variation although alterations also occur in other genes (Jason *et al.,* 2002). The genes that encode *gp120* mutate rapidly particularly in the V3 loop, the most immunogenic region of *gp120* (Spira *et al.,* 2003). The mutations that lead to the substitution of the positively charged amino acid at specific position in the V3 loop correlates with the syncytium inducing property of the virus (Holm-Hansen, 2000).

HIV-1 has several mechanisms of establishing genetic variation that include point mutation, deletion, insertion, duplication as well as recombination (Domingo, 1998). The main cause of this high variability is the recombination of heterogeneous genomes by co-infection of cells, and the high mismatch error rate of HIV reverse transcriptase enzyme coupled with lack or absence of proof reading capacity of error-prone reverse transcriptase that can switch between templates during proviral synthesis (Blackard *et al.,* 2002). It has been reported that HIV RT has an average error rate of 1/700 per nucleotide incorporation (Spira *et al.,* 2003) that results in approximately 10 genetic changes per replication cycle. This is due to lack of 3′-5′ exonuclease proof reading activity (Preston *et al.,* 1998; Bebenek *et al.,* 1989).

Recombination can mediate the repair of defective retroviral genomes (Boulerice *et al.,* 1991), can increase viral diversity, or can accelerate the spread of beneficial mutations among viral quasispecies (Termin *et al.,* 1991). The increased variation potential mediated by recombination confers on retroviruses the capability to respond rapidly to changing selective pressures, either immunological, (Ondoa *et al.,* 2001), or pharmacological (Moutouh *et al.,* 1996), through the prompt generation of the fittest variants possessing the adequate set of mutations to elude those pressures (Golovkina *et al.,* 1994).

2.2.1 Human Immunodeficiency Virus Type 1 group M genetic variability

As a result of the extensive genetic variation, HIV-1 can be divided into groups M (major), O (outlier) and N (non-m, non-o); (Najera *et al.,* 2002). The *env* proteins of *gp41* can show variations of up to 30-50%. The N subtype appears to be phylogenetically equidistant from M and O (Spira *et al.,* 2003). The M group is the most prevalent among the three groups and has nine subtypes, all of which originate from central Africa. Based on the differences in the sequences of genes that encode *gp120*, HIV-1 group M can be divided into nine clades A, B, C, D, F, G, H and K (Robertson *et al.,* 2000). Other HIV genetic loci may be used to determine HIV-1 subtypes, although the degree of variation differs according to the genomic region analysed (Jason *et al.,* 2002). The amino acid distance in the *env* gene between the subtypes in the major group have reached 25-35%, while in the *gag* gene it is about 15% (Thomson *et al.,* 2002; Takebe *et al.,* 2004).

olecular epidemiological studies have indicated the presence of diverse HIV-1 subtypes and unique recombinants (Morison *et al.,* 2001). There are currently 16 recognised circulating recombinant forms (CRFs) identified based on complete genome sequences derived from at least three epidemiologically unrelated individuals (Kuiken *et al.,* 2002). Forms known as unique recombinant forms (URFs) have not shown any evidence of epidemic spread and are thought to arise due to secondary recombination of CRFs (Thomson *et al.,* 2002). Currently there are 30 of them (McCutchan, 2000).

Generation of recombinant retroviruses requires that two viruses infect a single cell, either simultaneously, by single transmission event or sequentially, in multiple transmission events. In HIV-1, recombination can occur between different strains of the same subtype (intersubtype recombination), or different groups (intergroup recombination) (Najera *et al.,* 2002). The cumulative picture emerging from these studies indicates that HIV-1 recombinant forms are much more prevalent, geographically spread, and diverse in the global pandemic than previously known. This applies to not only circulating recombinant forms (recombinant forms identified in at least three epidemiologically-linked individuals) but also to the URFs (recombinant forms found in a single individual or in a single epidemiologically-linked cluster; Najera *et al.,* 2002).

2.2.2 Molecular Epidemiology of HIV

The distribution of HIV subtypes around the world differs in different parts of the world. Before 1992, HIV-1 strains were classified on the basis of their geographic origin into two subgroups, North American and African variants (Myers, 1994). The strains are now classified according to their molecular differences (Salminen *et al.,* 1995). In Africa HIV-1 subtype diversity is the highest in the world. All subtypes have been found although the A and C subtypes seem to be the most prevalent. The high diversity is probably a consequence of the virus originating from Africa (Janssens *et al.,* 1997).

There have been several molecular epidemiological studies of HIV-1 subtypes in Kenya. Most of these studies have centered on the analysis of partial sequences within the *gp 120* coding region of *env*. This involved the analysis of the *C2-V3* region (Janssens *et al.*, 1994; Robbins *et al.*, 1999). These studies found that the majority of the sequences analysed were subtype A (71-87%), with significant components of subtype D (7-29%) and subtype C (7-17%). In a study carried out by Dowling *et al.* (2002), where 41 near full-length sequences were analysed, a high proportion of recombinants was seen (40%). These full-length sequences also showed that there was a near absence of pure subtype C and D strains, with only 2.4% of each. Almost all the non-recombinants were subtype A, which comprised 56% of all strains (Dowling *et al.*, 2002).

2.3 The Host Immune Response to HIV Infection

The successful development of any protective vaccine requires knowledge of the immune correlates of protection (Letvin, 1998). Rational targets for vaccine elicited immune responses can only be established with an understanding of the immune responses that provide protection against infection by a pathogen (Zolla-Pazner, 2004). Defining such immune correlates of protection against HIV has proven extraordinarily difficult. Although powerful non-human primate models for HIV infection exist, impressive protection against AIDS virus challenges in these models has been difficult to achieve safe, vaccine strategies (Letvin, 1998).

Results of studies in non- human primates do not state the degree of confidence with which humoral or cellular immunity is needed to achieve protection against exposure to HIV (Miller *et al.,* 1994; Kuller *et al.,* 1994). The few reported cohorts of multiple exposed HIV sero-negative individuals are a focus of intense interest in the AIDS vaccine research community. Small cohorts of commercial sex workers in Nairobi and the Gambia, some infants born of HIV + positive mothers, some medical workers and various other small groups of people exposed to the virus show some resistance to infection (Rowland –Jones and McMichael, 1995; Rowland –Jones *et al.,* 1998). The exact mechanism of protection has not been fully established, but cell-mediated responses and possibly secretory antibodies are believed to be important (Walker *et al*., 1991; Walker 1993; Rowland –Jones *et al.*, 1999). The careful evaluation of multiply exposed, uninfected individuals may provide a means of defining the elusive immune correlates of protection. If the protection of these individuals against infection is immunologically mediated and the mechanism of this immune-mediated protection can be characterized, then it will be possible to define the type and level of immunity that must be elicited by vaccines to achieve protective immunity (Letvin, 1998).

2.3.1 Role of Innate Immunity in HIV Infection

Innate immunity is a non-specific first line of defense and is activated within hours of antigen contact and has no memory (Siegal and Spear, 2001). The role of innate immune response in HIV has not been well studied nor has its potential in vaccine immunity been explored (Landay and Heeney, 2001). The innate immune response is critical early in the host immune response before the adaptive response develops and plays a crucial role in helping to expand the adaptive immune response. It consists of both cellular and soluble components, which recognize pathogens by specific patterns displayed on their surface antigens (McMichael and Rowland-Jones, 2001) Components of the innate immune response include, dendritic cells, macrophages, and interferon gamma producing cells, NK cells granulocytes, and NK non cytotoxic T cells, chemokines, B1 cells, cytokines, and complement mannose binding lectins, defensins and acute phase proteins (Landay and Heeney, 2001).

The complement system appears to be involved in all stages of the HIV life cycle: the choice of the target cell, mechanism of entry into the cell, the activation of proviral DNA via signaling through complement receptors, viral budding and extra-cellular survival, all in different ways involve complement products (Sölder *et al.*, 1989). Human immunodeficiency virus activates the complement system even in the absence of specific antibodies. However, the role of complement in HIV infection has been largely underestimated, because the virus shows an intrinsic resistance to complement mediated lysis (Cooper, 1994). This is avoided by complement regulatory molecules which are included in the virus membrane upon budding from infected cells by decay acceleration factor (DAF/CD55) or are secondarily attached to HIV envelope glycoproteins such as factor H (Heribert *et al.*, 1997). Human Immunodeficiency virus takes advantage of human complement activation for enhancement of infectivity, for follicular localization and for broadening its target cell range (Dierich, 1992). At the same time, it displays an intrinsic resistance against the lytic action of human complement (Heibert *et al.*, 1997).

Macrophages express co-stimulatory molecules such as CD80 and CD86, which potentiate the adaptive immune response (Meylan *et al.*, 1993). They are derived from monocytes that enter the extra vascular tissue and are activated by cytokines, chemokines, lipopolysaccharide and mannose from pathogens (Montaner and Gordon, 1994). Macrophages produce cytokines and chemokines that inhibit the virus directly such as IL-1, IFNa/b, TNFa, nitric oxide, activate virus specific cellular responses (IL-1, IL-2) and recruit inflammatory cells (Cocci *et al.,* 1995).

Natural killer (NK) cells are a subset of peripheral blood leucocyte (PBL) that kills virus infected cells and tumours through an MHC-unrestricted mechanism and without prior sensitization (Herberman *et al*., 1986). Natural Killer cells can lyse a wide variety of virally infected cells either directly or through antibody dependent cellular cytotoxicity (ADCC; Herberman *et al*., 1986). They also effectively recognize and lyse infected cells in which the virus may have inhibited antigen presentation and MHC class I expression (Parker *et al*., 1995). They kill their targets through engagement of their Fc or complement receptors (Parker *et al*., 1995). They also produce Granulocytes-Macrophage Colony Stimulating Factor and like other cells of the innate immune system also produce IFNs and β-chemokines Type I IFN enhance the ability of NK cells to lyse infected target cells (Hu *et al*., 1995). Soluble factors in innate immunity bind to HIV directly or opsonise HIV infected cells to facilitate phagocytosis. Chemokines recruit cells to the sites of HIV infection and also block viral entry to CD4+ cells. They determine whether the immune response will be predominantly TH-1 or TH-2 type and can inhibit HIV replication directly (Hosein, 1995).

2.3.2 Role of the Adaptive Immune Response in HIV Infection

The major effectors of adaptive immune response to HIV and other viral infections include cytotoxic T cells (CTLs), helper T Cells, specific antibodies secreted by B cells (humoral and cellular immune mechanisms). Recognition of antigens in adaptive immunity is dependent on the T cell receptor and the Fab antigen binding regions of antibodies (Heeney *et al.*, 1998). These are able to recognize an extensive range of antigens through their enormous diversity achieved through comprehensive gene arrangement during B and T cell development (McMichael and Rowland-Jones, 2001).

T cell receptor diversity is determined by 3 hypervariable regions known as complementarity Determining Regions1-3 (CDR1-3; De Groot *et al.*, 2001). These regions interact to recognize peptide epitopes presented in the binding grooves of human leucocyte antigen (HLA molecules) which fall into two classes. Type 1 HLA molecules bind 8-12 amino acid peptide epitopes in their binding groove and present these to T cell receptors located on CD8+ T cells (Sette and Sydney, 1999). Type II HLA molecules bind longer peptides and present these to CD4+ T cells for recognition.

Preventive vaccines work through establishing immunologic memory for antigenic structures (epitopes) presented by the pathogen or by the infected cell (Letvin, 1998). Vaccine-induced immune response is induced prior to infection and can be recalled more rapidly than primary effector mechanisms. This involves the basic cellular elements of adaptive immunity, which includes B and T lymphocytes (McMichael and Rowland-Jones, 2001).

2.3.3 Role of CD4 Cells in Immunity to HIV-1 Infection

The primary effector mechanisms important for protection against viruses are antibodies produced by B cells and cytolytic activity that can influence differentiation, expansion and duration of T cell responses (McMichael and Rowland-Jones, 2001). The CD4+ T cells are the heart of the immune system, responsible for co-ordination of the immune response. This is regulated through the coordinated multi-functional roles, primarily executed through the coordinated release of cytokines and direct cell surface interactions involving co-stimulatory molecules (Rosenberg *et al.*, 1997)

T cells recognize virus-infected cells by specific interactions between the T cell receptor and 8-10 amino acid peptides processed from viral antigens and presented in the context of MHC molecules (Lewis *et al.*, 1994). Therefore T cells clear viruses effectively after infection has occurred .The recognition is restricted by MHC molecule, particular epitope recognition by a given individual will depend on the set of inherited alleles encoding the MHC molecules (Elliott *et al.*, 1994; Jorgensen *et al.*, 1992).

The hierarchy of recognition or epitope dominance may vary among individuals who share MHC halotypes (Chaturvedi *et al.*, 1996; Constant *et al.*, 1995). Hence epitope repertoire in a vaccine will need to have enough breadth to encompass all the relevant MHC halotypes of potential vaccines. The need for CD4+ T lymphocytes to initiate the adaptive immune response presents a dilemma since these cells are the major targets for HIV-1 infection. CD4+ T cell may have some capacity for lysis of HIV-infected cells (Silesian *et al.*, 1988) and production of anti-viral cytokines. The major role is in shaping immune response by establishing a microenvironment with a particular cytokine profile such as IL-12, IL-2 and IFN-γ, which provides protection, than induction of TH-2 cytokines such as IL-4, IL-5 and IL-13. CD4+ T cells, through expression of IL-2, provide help for maturation of CD8-T cells into CTL effectors (Tsomides *et al.*, 1994).

Interferon gamma (IFN-γ) activates macrophages, increasing their capacity to kill intracellular pathogens (Stout and Bottomly, 1989). Interleukin -4 (IL-4) and IL-6 as well as co-stimulatory molecules for example CD40 ligand are essential for the coordinated production of antibodies from B cells (Wilson *et al.*, 1998). CD4+ T cells are also of obvious importance, especially for influencing differentiation patterns and expansion of selected lymphocyte population, but their roles as direct effectors of virus clearance is not clear (Mosmann and Sad, 1996; Scott, 1993). CD4 T cells will be induced in the process of achieving the appropriate antibody and CD8+ CTL responses (Paliard *et al.*, 1988). Initial priming of vectors and the use of adjuvants other than alum in HIV vaccines (which promotes TH-2 responses) provides an advantage (Chaturvedi *et al.*, 1996). Human Immunodeficiency virus infection results in progressive loss of CD4+ T cells from the circulation as well as depletion of CD4+ T cells from total body stores. The virus can have direct and indirect pathogenic effects on both mature CD4+ T cells and on the progenitors (McCune, 2001).

During HIV infection, TH-1 cytokines levels remain high in the asymptomatic period. This group of cytokines includes IFN-γ, IL-2 and IL-12. The high levels of TH-1 cytokines may itself help to suppress levels of TH-2 cytokines for example IFN-γ appears to inhibit the production of Th-2 cytokines (Openshaw *et al.*, 1995). For those people whom HIV progresses to AIDS, the shift begins during this asymptomatic period (Sean Hosein, 1995). T Helper-1 cytokines start to fall and the level of TH-2 cytokines (IL-4, 5, 6 and 10) starts to rise. Elevated levels of IL-4 (TH-2 cytokine) can stimulate production of Th-2 cytokines including (IL-4 itself; Swain *et al.*, 1990; LeGros *et al.*, 1990).

2.3.4 Role of CD8+ Cytotoxic T Lymphocytes (CTL)

Control of initial viremia associated with primary infection temporarily correlates with the appearance of CD8+ CTL, and mutation in specific CTL epitopes can be detected in residual virus population (Borrow *et al.*, 1997; Price *et al.*, 1997; McMichael and Phillips, 1997). The appearance of HIV-specific CTLs after primary infection has been shown to be concurrent with a dramatic fall in viral load (Koup *et al.*, 1994). Human Immunodeficiency virus specific CTL activity has been demonstrated in a small subset of uninfected seronegative commercial sex workers in Gambia and Kenya. This suggests that transient infection may have occurred, inducing CD8+ CTL mediated protective immunity (Forke *et al.,* 1996; Rowland-Jones *et al.*, 1995).

Neutralizing antibodies and cytolytic T cells are the major effectors of anti-viral immunity. CD8 T cells are the principal effector mechanisms of adaptive immunity to clear virus infected cells (Barńey, 2000). This has been demonstrated exhaustively in Sinai virus, influenza virus, respiratory syncytial virus, herpes simplex (Lu *et al.*, 1980;

Ada *et al.*, 1981; Graham *et al.*, 1991; Sethi *et al.*, 1983). Cytotoxic T lymphocyte (CTL) effector function falls into two pathways: the perforin /granzyme pathway and the Fas/Fas ligand pathway (Williams and Engelhard, 1997).

The CD8+ T-lymphocytes recognize virus-infected cells through a cognate interaction between the T cell receptor and a processed peptide epitope presented in the groove of a MHC Class 1 molecule. The lysis of infected cells occurs through the production and secretion of perforin and granzyme (essential for calcium dependent direct cell lysis) that penetrate the target cell membrane and induces apoptosis (Williams and Engelhard., 1997). Fas ligand (Fasl; which mediates calcium-independent cell lysis) is also unregulated on the activated CD8+ T-cell, which can bind Fas (a type 1 transmembrane protein) on the target cell and induces apoptosis through other pathways. Cytotoxic T lymphocytes also produce cytokines with antiviral properties such as IFN-γ and TNF-γ inhibiting viral replication (Borrow *et al.*, 1997).

There is increasing evidence that a substantial proportion of people with documented HIV exposure who remain uninfected generate a range of immune responses to the virus, which include TH-1 responses, CTL (Clerici *et al.*, 1994) and mucosal Ig-A secretion (Rowland-Jones *et al.*, 1995) while remaining negative for plasma HIV antibodies. In long-term progressors (LTP), some of these individuals are infected with virus isolates that replicate poorly (Deacons *et al.*, 1995; Kirchhoff *et al.*, 1995), however others are infected with the virus isolates that have normal replication capacity, but have maintained a strong and broad set of humoral and cellular HIV-specific immune response that appears to be responsible for delayed disease progression. This has been associated with HIV-specific CD4+T cell proliferation (Rosenberg *et al.*, 1997) and a strong CD8 + CTL activity against multiple epitopes (Rowland-Jones *et al.,* 1999; Dyer *et al.*, 1999).

Vaccines work through induction of CTL responses. In Macaques immunized with recombinant Modified Vaccinia Ankara (MVA) prior to challenge with SIV, showed that vaccination did not prevent infection and the CTL cell response was associated with delayed disease progression (Hirsch *et al.*, 1996). Approaches to optimize the CD8 CTL such as the addition of an IL-2 adjuvant to a recombinant DNA vaccine regiment, nearly complete control of subsequent Simian Human Immunodeficiency virus (SHIV) infection can be achieved (Barouch *et al.*, 2000).

2.3.5 Role of Antibodies in the Control of HIV Infection

Although cell mediated immunity is crucial for controlling and eradicating infection by many viruses, antibodies are pivotal in preventing or modulating infection (Robbins *et al.*, 1995; Hilleman, 2001). Antibodies are the only components of adaptive immune response that can neutralize a virus particle prior to infection of a cell, unlike T cells, which only recognize virus in the context of an already infected cell, and require a few days for activation and expansion of memory populations to respond (Burton *et al.*, 2000). During infection there is a strong antibody response to envelope proteins, but this response is ineffective against mature virions (Mo *et al.*, 1997). Several studies have shown that antibody alone is not enough to offer prophylactic or therapeutic benefit against HIV when infused (Cavacini *et al.*, 1998; Fletcher *et al.*, 2000). Although in macaques antibody infusion has acted to prevent SHIV transmission via the mucosa (Mascola *et al*., 2000).

Antibodies generated during HIV infection are possibly directed towards viral fragments whose configuration may differ from whole virus (Mo *et al*., 1997). One obstacle to the development of an effective HIV-1 vaccine has been the difficulty of inducing broad reactive, potent antibodies with protective functions (Fletcher *et al*., 2000). During the early phase of infection, when the level of antibodies induced by active immunization or administered by passive immunization is low, an effective neutralizing antibody response is a critical component of vaccine-induced immunity because it can reduce the size of the infecting innoculum and neutralize or eliminate virions during the first rounds of replication (Zolla-Pazner, 2004; Figure 2.3). This provides sufficient transient protection so that the cellular arm of the immune response can respond with proliferation and deployment of effector T cells that are required to eliminate virus-infected cells (Haigwood *et al*., 1996; Hilleman, 2001). Many passive immunization experiments in experimental models have, established that antibodies can provide sterilizing immunity against HIV-1 (Emini, 1995; Mascola *et al*., 2000). It has been proven in non-human primate models of lentivirus infection that sufficient levels of neutralizing antibody (nAb) can prevent infection (Emini *et al*., 1992).

Recently, passive prophylaxis using HIV immune globulins combined with monoclonal antibodies (mAbs) has protected macaques from virginal challenge with SHIV (Mascola *et al*., 1999) and a mixture of three neutralizing monoclonal IgG1 antibodies given to pregnant macaques has protected infants from SHIV oral challenge (Baba *et al.*, 2000). These studies provided convincing data in all experimental models tested that neutralizing antibodies, with appropriate specificity, when present in sufficient concentration, are a correlate of immune protection (Mascola *et al.*, 2000).

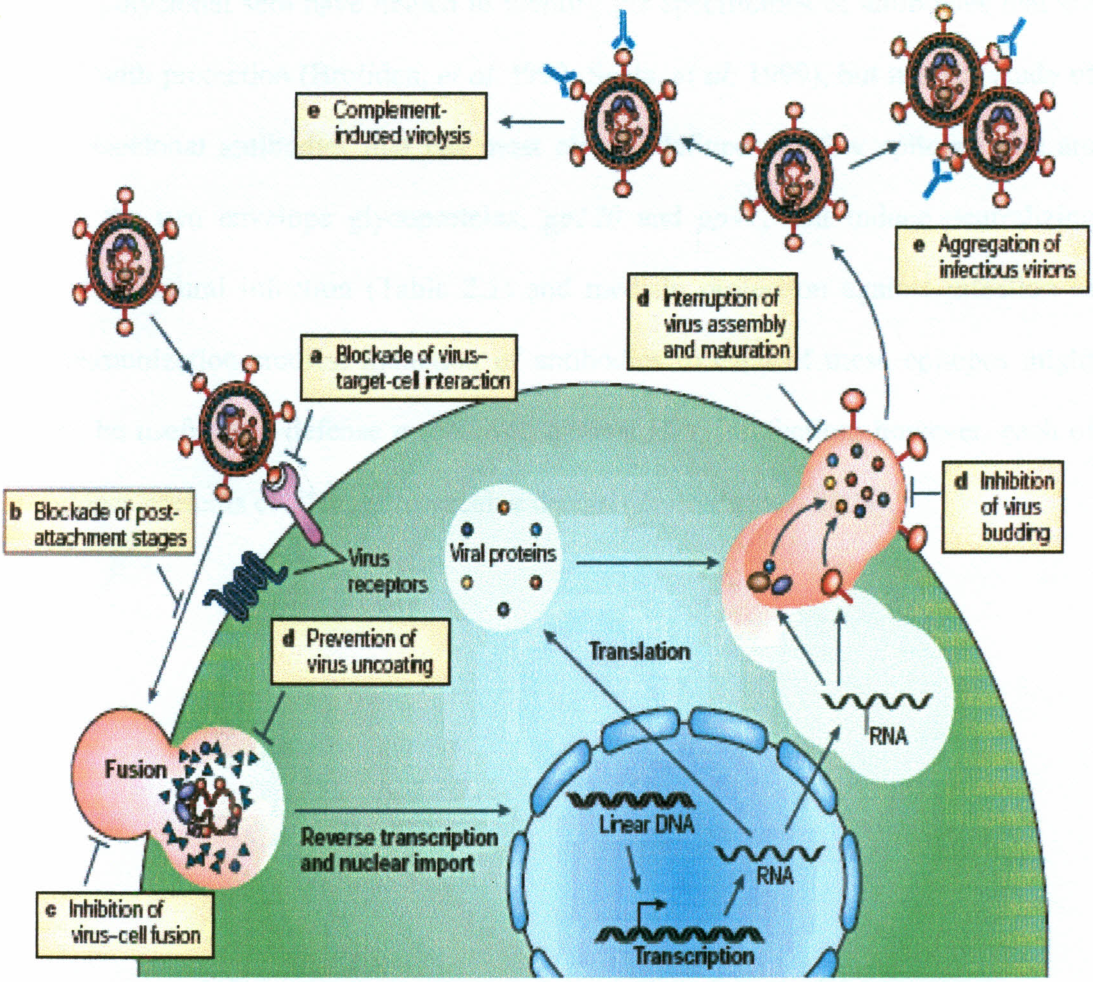


Figure 2.3: Steps at Which Antibodies Can Potentially Interfere With HIV-1 Virus Replication (Adopted From Susan Zolla-Pazner, 2004).

2.4 Identification of Protective Epitopes

The most consistent correlate of immune protection against viral infections is the presence of neutralizing antibodies (Dimmock, 1993; Robbins *et al.,* 1995). Emphasis has been placed on identifying antibodies that neutralize HIV-1. The virus induces a polyclonal antibody response to a wide array of epitopes on different viral proteins (Dowbenko, *et al.* 1988; Goudsmit, 1988).

Studies of polyclonal sera have helped to identify the specificities of antibodies that are associated with protection (Broliden, *et al.* I992; Scala, *et al.* 1999), but it is the study of human monoclonal antibodies that has most clearly defined the few epitopes that are located in the two envelope glycoproteins, *gp120* and *gp41*, that induce neutralizing antibodies in natural infection (Table 2.1) and mediate protection against infection in passive immunization studies. Induction of antibodies to each of these epitopes might ultimately be useful as a defense mechanism against HIV-1 infection; however, each of these epitopes presents challenges to vaccine design (Zolla-Pazner, 2004).

Table 2.1: Epitopes of the HIV-1 envelope glycoproteins that induces neutralizing antibodies. Adapted from Zolla-Pazner (2004).

Epitope	**Representative human monoclonal antibody**	**Characteristics**
Cluster 1 of *gp 41*	Clone 3, 246-D	Highly immunogenic epitope, but clone 3 is the only one of many monoclonal antibody specific for this epitope
Transmembrane-proximal region of *gp 41*	2F5, 4E10, Z13	Poorly immunogenic but antibodies to this region are broadly neutralizing
CD4 binding domain *of gp 120*	Clone 3,264-D	Highly immunogenic, but IgG1b12 is the only monoclonal antibody specific for this epitope that has broad neutralizing activity
CD4-induced epitope of *gp 120*	IgG1b12, 559/64D, 15e	Only antigen binding fragments of antibody specific for this epitope are neutralizing. Intact IgG molecules specific for this epitope are neutralizing
1α mannose residue of gp 120	17b, 48d	Poorly immunogenic but at least one monoclonal antibody to this region is broadly neutralizing.
V2 loop of gp 120	2G12	Highly immunogenic but antibodies to these epitopes are isolate specific
V3 loop of *gp 120*	447/52-D, 19b, 2182	Highly immunogenic, but antibody specificity broadens only after antigenic stimulation

2.4.1 Regions of HIV-1 *gp*41 and *gp*120 that are Crucial to Virus Function and Immune Protection

2.4.1.1 Envelope Glycoprotein 41 (*gp* 41)

This transmembrane glycoprotein is found as a homotrimeric complex in the envelope of the virus; it interacts noncovalently with *gp*120 on the exterior of the virus particle. The immunodominant region (includes *gp*41 epitope cluster I): induces high levels of antibodies, most of which are not neutralizing but might mediate other functions, such as antibody-dependent cell-mediated cytotoxicity (ADCC) and aggregation of and complement deposition on virus particles (Cotropia *et al*., 1996). The transmembrane-proximal region is a poorly exposed region on the surface of the virus and is thought to be a transitional epitope, exposed for a brief period of time during the conformational changes that occur in *gp*41 that lead to the fusion of virus and cell membranes (Zwick *et al*., 2001; Muster *et al*., 1997). The amino-terminal and carboxy-terminal heptad repeat regions leucine zippers, which are involved in the formation of the coiled-coil form of *gp*41 after conformational changes are induced in *gp*120 by its interaction with CD4 and chemokine receptors (Xiang *et al*., 2002). The fusion peptide: the amino-terminal region of *gp*41, which is exposed after formation of the coiled-coil form. This region is inserted into the membrane of the target cell, resulting in the fusion of virus and cell membranes (Ho *et al*., 1991).

2.4.1.2 Envelope Glycoprotein 120 (*gp* 120)

Different regions of *gp*120 interact with CD4 and chemokine receptors found on the surface of target cells (Thali *et al*., 1993). These interactions lead to a series of conformational changes in *gp120,* and subsequently allow the conformational changes in *gp*41 that lead to formation of the coiled-coil form and exposure of the fusion peptide (Sullivan *et al*., 1998).

The CD4-binding domain (CD4bd): composed of several parts of the *gp*120 molecule, forming a binding pocket into which a region of CD4 fits (Sullivan *et al.*, 1998). Bridging sheet: composed of four anti-parallel β-strands from the V1/V2 stem and the C4 regions of *gp*120 is involved in the binding of *gp*120 to chemokine receptors, triggering subsequent conformational changes in *gp*120 and *gp*41. This region forms or is exposed after binding of *gp*120 to CD4, is known as the 'CD4-induced epitope' (Moulard *et al.*, 2002; Feng *et al.*, 1996).

The V2 loop of the gp120 is a highly variable region that is proximal to the CD4bd and is part of the bridging sheet. The V2 loop (together with the V1 loop) seems to shield partially the CD4bd, the bridging sheet and part of the V3 loop until conformational changes in *gp*120 are induced by CD4 (Gorny *et al.*, 1991). The V3 loop: a semi-conserved region of *gp*120 that is structurally constrained by its requisite participation in virus infectivity. The V3 loop interacts with chemokine receptors on the surface of target cells (Zolla-Pazner, 2004).

2.5 Concepts and Designs of Vaccine Development

2.5.1 Ideal Characteristics of an AIDS Vaccine

An ideal HIV preventive vaccine should protect against any route of infection, induce protection against infection with diverse viral isolates, (preventing the need for many isolate- specific vaccines), provide long lasting protection, (require only a single dose administration) and be stable and easy to administer, (facilitating mass immunization campaigns in the developing countries with minimal infrastructure). It should also be effective regardless of the nutritional status, low cost and have excellent safety profiles, with minimal risk of adverse reactions even in unscreened populations (Zolla-Pazner, 2004).

Clinical trials are conducted in three main phases with slight variations from product to product. In phase I trials, there are few volunteers and the focus is on safety issues and immunogenicity. The side effects of the product are evaluated. In phase II trials, the vaccine is tested to see if it stimulates an appropriate immune response and how this can be optimized. Safety and efficacy issues are also evaluated and compared with placebo. In phase III, candidate vaccine, now seen to be safe and immunogenic is tested in many volunteers to see if it is effective (Hanke and McMichael, 2000).

2.5.2 The Status of HIV-1 Vaccines

In view of the complexity of the HIV virus, the inadequacy of cell mediated and antibody response and accumulating data from primate vaccination, it may prove impossible to develop a vaccine that provides sterilizing immunity once HIV infection has occurred. However, the need for a prophylactic or therapeutic HIV vaccine grows daily. A regimen that attenuates infection or postpones the onset of AIDS for the lifetime of the individual may be feasible (Hanke and McMichael, 2000). Inspite of an overall desire to develop a vaccine as soon as possible, the research fraternity seems nowhere near therapeutic or prophylactic vaccine (Zolla-Pazner, 2004). This is because major stakeholders such as pharmaceutical companies are only partially engaged possibly because a successful vaccine could reduce sales of antiretroviral, but also because of the high cost of research (Hanke *et al.*, 2002). It is also difficult to get large enough primate populations for laboratory experiments since bodies that regulate clinical trials seem overly cautious and are slowing trials considerably (Hanke *et al.*, 2002). Poor infrastructure, few resources, unusual cultural and religious believes on HIV/AIDS hinder vaccine trials in general (Ramsay *et al.*, 2002).

Given this degree of diversity, it is widely believed that a vaccine based on a single strain or subtype of HIV-1 will not be successful against the large spectrum of globally circulating HIV-1 variants (Mahmudul and Mahbub, 2003). Most vaccine formulations in clinical trials contain immunogens derived exclusively from subtype B viruses, the predominant genotype in the United States and Europe (Table 2.2). Little emphasis has been placed on the development of candidate vaccines on non-B viruses that cause the vast majority of HIV-1 infection in developing countries (Gao *et al.*, 1996). It is clear that phase I trial of HIV vaccines should be conducted for as many vaccine candidates as possible with the view to selecting the most promising one for phase II and III. Only one gp120 subunit vaccine initiative has entered phase III clinical trial so far (Mwau and McMichael, 2003).

The challenge lies in increasing the number and types of candidate vaccine, bolstering the participation of major stakeholders sourcing for adequate funding and animal models and stimulating a healthy demand for HIV vaccines and understanding the immune response to HIV infection for an eventual HIV vaccine to be feasible. In order to improve the immunogenicity and efficacy, better vaccination strategies such as prime-boost strategies (Hanke and McMichael, 2000; Ramsay *et al.*, 2002), gene gun delivery (Hanke and McMichael 2000) and use of multiple vectors should be utilized.

Table 2.2 HIV-1 Clade Composition of HIV-1 Vaccine Candidates

Vaccine type	**Company group**	**Prime HIV strain**	**Boost HIV-1 strain**
Recombinant protein (subunit)	Vaxgen	BB	-
Recombinant protein (subunit)	Vaxgen	BE	-
DNA recombinant protein	Chiron	B	B
Canarypox recombinant protein Boost	Aventis Pasteur	B	B
DNA	Mecrk	B	B
DNA MVA	Oxford AIDS vaccine initiative	A	A
Replicon	Alphvac	C	C

2.6 Approaches to HIV Vaccine Development

2.6.1 Recombinant Subunit Vaccines

Recombinant protein immunogen (subunit vaccines) are subunits of HIV such as *Env*, *Tat* or *P*24. Subunit vaccines for *gp* 120 and *gp*160 have already been developed and assessed. These vaccines are relatively safe (Baba *et al.,* 1995). The antigenicity of a subunit protein is affected by the conformation and the specific cell type used for its production (Keefer *et al.*, 1994). Due to the high degree of glycosylation of the HIV-1 env, the amount and pattern of glycosylation of the recombinant envelope has critical effect on its ability to induce neutralizing antibodies (Burton, 1997). Subunit vaccines do not induce endogenous synthesis of viral proteins in APCs hence do not induce CD8+CTL (Keefer *et al.*, 1994). Neutralizing antibodies generated by subunit vaccines tend to be strain specific. One *gp 120* subunit vaccine on trial (Francis *et al.*, 1998) did not induce antibodies to neutralize field isolates and the volunteers in phase 1 and 11 became infected later after exposure tox HIV (Francis, 2001). Recombinant proteins that can raise serum antibody levels needed for protection are yet to be developed.

2.6.2 Live Recombinant Vaccines

Live recombinant vaccines present antigens in a nearly natural conformation; glycosylation and oligomerization (mimic antigen presentation that occurs during natural viral infection). They have the potential to cause disease especially in immuno-compromised hosts. This can be partly overcome by the use of highly attenuated strains, even though this reduces immunogenicity (Baba *et al.*, 1995). A number of vectors such as Vaccinia and Canary Pox have the ability to carry large segments of DNA. The presence of promoter sequences before the inserts ensures production of antigens in significant quantities. The commonest viruses employed to carry HIV genes include Poxviruses (Targlia *et al.*, 1992; Abimiku *et al.*, 1995), Semliki forest virus (Mossman and Sad, 1996), Adenoviruses (Natuk *et al.*, 1993; Davis *et al.*, 2000; Poliovirus and sindbis virus (Villacres *et al.*, 2000). Some poxvirus vectors such as MVA have undergone deletions in their genome that make them no longer immunosuppressive or pathogenic. The Oxford/Nairobi HIVA vaccine initiative employed MVA as one of the vectors for HIVA vaccine (Hanke and McMichael, 2000).

2.6.3 Whole Inactivated HIV-1 vaccines

This approach has been effective in vaccine formulations such as Polio (Salk vaccine) and influenza vaccines (Villacres *et al*., 2000). A therapeutic trial of one of these vaccines, REUNE (Moss *et al*., 2002) has shown some evidence of T helper cell response augmentation. The potential risk associated with incomplete inactivation of the virus stock, have raised concern for such an approval with HIV. Excessive inactivation would lead to the disruption of the structure of potential neutralizing antibodies (Chakrabarti *et al*., 1996).

2.6.4 Live Attenuated Vaccines

Live-attenuated vaccines that express majority of HIV proteins have been able to elicite the most complete, long lasting immunity and have proved to be the best immunogens in primate experiments (Desrosiers, 2004). Vaccination with attenuated SIV strains with defective *nef* genes has been the most effective approach tested (Wyand *et al*., 1999). Such genetically altered viruses have been used as vaccines to prevent polio, measles and chicken pox in humans (Letvin, 1998). They simulate the responses that occur in natural HIV infection. The use of such vaccines in humans is hampered by overwhelming safety concerns (Putkenen *et al*., 1991; Wyand *et al*., 1999)

Heavily attenuated SIV that protected adult macaques from SIV infection (Wyand *et al*., 1999) produced AIDS in neonatal macaques (Baba *et al*., 1995). A few of the protected adult macaques eventually developed AIDS (Daniel *et al*., 1992). The degree of attenuation inversely correlates with immunogenicity (Johnson *et al.*, 1999). Approaches to increase the safety of these constructs includes causing further attenuation by deleting several areas of accessory genes (Desrosiers, 2004) or incorporating a suicide gene thus conferring susceptibility of the virus to antiviral agents (Chakrabarti *et al.*, 1996). Human immunodeficiency virus type 1 mutates extremely rapidly, therefore it was suggested that accumulating genetic alterations in a non pathogenic virus used as vaccine might lead to the virus eventually regaining its pathogenic potential (Letvin, 1998).

2.6.5 Deoxyribonucleic Acid (DNA) Vaccines

Deoxyribonucleic Acid vaccines consist of a gene or genes of interest vectored in plasmids (Amara *et al.*, 2001). The plasmid is designed to enable high levels of expression in eukaryotic cells. Strong promoters for optimal expression of mammalian cells as well as polyadenylation sequence to stabilize mRNA transcripts are also included (Amara *et al.*, 2001). The efficacy of DNA vaccines in trials in macaques has already been demonstrated; usually preventing SIV induced AIDS but not providing sterilizing immunity (Barouch *et al.*, 1998; Barouch *et al.*, 2000; Amara *et al.*, 2001).

Successful induction of both humoral and cellular immune responses as well as mucosal immune responses has been achieved in primates. DNA vaccines are weak and can be improved by a number of strategies such as cytokine genes in DNA constructs, co-immunization with chemokines, use of cytidine-phosphate –guanosine (CpG) motifs, prime-boost vaccination and the use of gene-gun technology (Letvin, 1998). Successive immunization with DNA and modified Vaccinia virus Ankara (MVA)-based vaccines expressing common immunogen is a potent way of inducing CD8+CTL (Amara *et al.*, 2002; Hanke *et al.*, 2002; Wee *et al.*, 2002).

Encouraged by the immunogenicity of this approach in non-human primates, a DNA/MVA-based vaccine was designed and constructed in Oxford (Hanke and McMichael, 2000) for clinical trials in humans. This was the first HIV-1 clade A- derived vaccine to be tested in humans. The immunogen was derived from consensus HIV clade A *gag* p24/p17 sequences and a string of 25 clade A-derived CTL epitopes. It did not contain envelope genes and therefore focused on inducing cell-mediated immune responses. It was designed for phase III efficacy trials in high-risk cohorts in Kenya, where 70% of infections are caused by HIV clade A (Neilson *et al.*, 1999).

2.7 Need for a HIV Vaccine

HIV infection continues to spread at an alarming rate. The continued spread of the HIV pandemic in both industrialized countries and the developing countries provides compelling evidence for the need of an effective AIDS vaccine. HIV transmission is in theory largely preventable, but without the development of an effective vaccine, it isn't (Hanke and McMichael, 2000). HIV will continue to infect millions throughout the world. Education has been slow to make an impact and highly active anti-retrovirus therapy is, for majority of people, too expensive and complex and in any case fails to clear the virus from the body (Letvin, 1998). Though the Kenyan government has provided the antiretroviral drugs for free, accessibility is still a problem.

While programs to reduce HIV transmission of HIV have achieved some success in both developed and developing countries, it is unlikely that the wide spread application of these programs will be able to achieve sustainable decrease in HIV transmission. Although the advent of highly effective antiretroviral therapy has resulted in significant increase in survival of HIV-infected individuals, the impact of combination antiretroviral therapy is confined to the industrialized world, which at present constitutes less than 10% of the world HIV infected population (McMichael and Hanke, 2000). Moreover, whereas the need for HIV vaccines is urgent and overwhelming, research and development of HIV vaccine has been hampered significantly by an overall lukewarm approach by the world community especially Pharmaceutical companies who are major stakeholders and only partially engaged because the profits they reap from the sale of antiretroviral drugs are tremendous but also because cost of research is high and continous mutation of the virus. The challenge lies in increasing the number and types of candidate vaccines, bolstering the participation of major stakeholders, sourcing of adequate funding and animal models and stimulating a healthy demand for HIV vaccines (Hanke and McMichael, 2000).

2.8 Approaches to enhance immunogenicity of HIV vaccines

In order to improve the immunogenicity and efficacy, vaccination strategies are becoming increasingly sophisticated. Several strategies have shown merit such as prime-boost strategies (Hanke *et al.*, 2002; Hanke and McMichael, 2000; Ramsay *et al.*, 2002), gene gun delivery (Hanke and McMichael 2000) and use of multiple vectors.

Improving vaccines to generate more robust and long lasting T cell response is important for both humoral and cell mediated immunity (Rook *et al.*, 1995). In humans DNA vaccines are expected to at best abort HIV spread by rapid expansion of memory HIV specific CD8+ population soon after exposure to HIV. However DNA vaccines expressing envelope proteins gave weak and ineffective antibody response (Rook *et al.*, 1995). This can be improved by a number of strategies such as including cytokine genes (IL-2) in DNA constructs, co-immunization with chemokines, use of cytidine-phosphate –guanosine (CpG) motifs, prime boost vaccination and use of gene gun technology (Hanke *et al.*, 2002).

Interleukin 12 (IL-12) is a recently characterized cytokine that may play a pivotal role in immuno-modulation such as the addition of IL-12 to an alum-adsorbed HIV-1 gp120 vaccine elicited type 1 (Th1) cytokines and IgG2 and IgG3 antibody responses in mice. (Luis *et al.*, 1994; Bliss *et al.*, 1996; Rook *et al.*, 1995). The same vaccine without IL-12 induced type 2 (Th2) cytokines and IgG1 antibody responses (Jankovic *et al.*, 1997).

Prime boost strategy involves 'priming' the immune system with one vaccine such as a live vector vaccine (a relatively non-virulent bacterium or non –HIV virus that has been genetically engineered to contain one or more synthetic HIV genes) and then 'boost' the subsequent immune response with a different vaccine such as gp 120 subunit recombinant vaccine (Hanke *et al.*, 2002). For example, DNA/MVA based HIV vaccine candidate was designed and constructed in Oxford for clinical trial in humans, used prime boost strategy and went through phase I and II for trial to prove safety and immunogenicity (Hanke *et al.*, 2002). Several experimental recombinant live vector vaccines made from canary pox and modified Vaccinia virus Ankara (MVA) have been engineered to carry foreign HIV genes in to the body hence stimulate production of protective antibodies and T cells (Rook *et al.*, 1995).

Formulation of vaccines with potent adjuvants is an attractive approach for improving the performance of vaccines composed of subunit antigens. Development of safe and effective vaccines composed of subunit antigens will require the ability to selectively drive appropriate protective immune responses to them. The use of immunologic adjuvants to enhance and direct immune responses to subunit vaccines is a critical component of a rational vaccine design (1997; Putkonen *et al.*, 1991; Kahn *et al.*, 1994). Presently, aluminum salt-based adjuvants incorporated into HIV vaccines continue to be the only immunologic adjuvants used in U.S.-licensed vaccines. Certain novel adjuvants such as purified saponins, Immunostimulatory complexes (ISCOMS), and liposomes have been shown to greatly improve the induction of MHC class-I-restricted CD8+ CTL responses over those induced by the same antigen given alone or in combination with standard alum adjuvants (Takahashi *et al.*, 1990).

2.9 Importance of Clades in HIV Vaccine Development

Human Immunodeficiency Virus continually evolves because of genetic mutation and recombination. Thus, there is need to account for strain variation within individuals and among populations when developing HIV vaccines (McMichael *et al.*, 2002). Whenever a drug or immune response destroys one variant, a distinct but related resistant variant can emerge. Any of these changes may yield a virus that can escape identification and attack by the immune system. Given that a preventive HIV vaccine will need to generate immune responses that protect uninfected individual from all HIV subtypes and recombinant forms, identification of conserved genes common to most or all subtypes and use as vaccine candidates may evoke broad- based immunity (McMichael *et al.*, 2002).

Clades differ by 5-17% in their amino acids (wang *et al.,* 1999) and therefore there are good reasons for candidate vaccines to match regional clades (McMichael *et al.*, 2002). The consequence of this amino acid differences is that the recognition of epitopes among clades is affected by two-thirds (Rowland-Jones *et al.*, 1998 and a phenomena such as antagonism enhanced (Montefiori *et al.*, 2001). For example the DNA vaccine (HIVA) developed in oxford for use in Kenya was based on HIV-1 clade, which is the commonest circulating clade in East and Central Africa (Hanke *et al.*, 2002) but was not efficacious enough to cover all the circulating strains.

2.10 Diversity Consideration in HIV-1 Vaccine Selection

Genetic variation of HIV-1 poses a major obstacle for the AIDS vaccine development. The divergent patterns of the AIDS epidemic in different geographic areas may be an important consideration for the design and testing of HIV vaccines (Hanke and McMichael, 2000).

Currently, candidate vaccines are derived from isolates, with the hope that they will be sufficiently cross-reactive to protect against circulating viruses (Novitsky *et al.*, 2002). This may be overly optimistic, given that strains belonging to the same subtype can differ by up to 20% in their envelope amino acid and between subtypes, distances can sour up to 35% (Korber *et al.*, 2002). The current scale of HIV-1 pandemic worldwide makes action imperative to design vaccine strategies to contend with extraordinary diversity of HIV (Gao *et al.*, 2005)

An effective vaccine may have to stimulate a range of host defenses including mucosal and innate immunity, neutralizing antibodies and cell-mediated responses (Hanke and McMichael, 2000). However, determining the level of protection that vaccine-induced CTLs can confer against HIV exposure and in already infected individuals on anti-retrovirus therapy will be possible only through development of strategies that reliably elicit strong and durable CTL responses in humans (Novitsky *et al.*, 2002).

Choosing the 'right' vaccine candidate by employing a particular viral strain, clone or isolate is still the current approach to HIV vaccine design. Higher homology between vaccine and circulating strain(s) results in a more efficacious vaccine (Novitsky *et al.*, 2002). For example, the need for frequent changes in the annual influenza vaccine put into perspective the implication of such diversity. Less than 2% of amino acid change can cause a failure in the cross-reactivity of the polyclonal response to the influenza vaccine and necessitate changing the vaccine strain (Korber *et al.*, 2002). Los Alamos National Laboratory has created an extensive global database of over 80,000 HIV-1 sequences linked to geographic and subtype information (www.hiv.lanl.gov). This database provides a framework for a reasoned selection of vaccine candidates for testing (Korber *et al.*, 2001)

2.11 Vaccine Strategies That Contend With HIV Diversity

There are a number of approaches to selecting vaccine strains that attempt to contend with the high levels of HIV sequence variation.

2.11.1 Isolate Based Vaccines

This approach is based on the use of isolates of a particular subtype, sometimes selected from a geographical region where the vaccine is intended for use, for example the development of HIV-1A trial in Kenya (HIVA vaccine; Hanke *et al.*, 2000), and subtype C- vaccine in Southern Africa (Novitsky *et al.*, 2002). The AIDS vaccine reagents were first developed from subtype B viruses, the dominant subtype circulating in the United States and Europe (Gaschen *et al.*, 2002).

There has been discussion of choosing a regional strain for vaccine for example an Indian strain for India and South African for South Africa (Goudsmit, 2001). Subtype sequences from Botswana and South Africa intermingle and there is no obvious choice of a single sequence most representative of the diversity in these regional samples. High levels of HIV-1 intersubtype in Southern Africa (Choudhury *et al.*, 2000; Novitsky *et al.*, 2000; Novitsky *et al.*, 1999) increase the magnitude of the challenge.

2.11.2. Artificial Sequence for Minimizing Diversity

An effective way to minimize the degree of sequence dissimilarity between a vaccine strain and contemporary circulating viruses is to create an artificial sequence by using: A consensus sequence based on the most common amino acid in each position of alignment (Novitsky *et al.*, 2002; Korber *et al.*, 2001). In constructing a consensus and ancestral sequence, hyper-variable regions are aligned by anchoring on glycosylation sites and only minimal common element spanning the region is retained. The consensus sequences are derived by careful characterization before use in a vaccine (Korber *et al.*, 2001).

Consensus sequences may be ideal for peptides used to explore the T cell immune response, as it would improve recognition. CTL escape mutations can rapidly predominate in the viral quasi species (Allen *et al.*, 2000) and may go undetected though the use of peptides based on isolates from later time points that have escaped the early responses may be useful (McMichael and Phillips., 1997).

A model of the most common ancestral sequence of an appropriate lineage can be constructed from a phylogenetic tree such as by means of maximum likelihood. Such sequences are central and similar to the currently circulating strains of interest and may have enhanced potential for eliciting cross-reactive response (Korber *et al.*, 2002).

2.11.3 Conserved and Ancestral Sequence Conserved CTL Epitopes

Experimentally defined CTL epitopes in the HIV database cluster more densely in conserved regions of the HIV protein. The epitopes in the database have primarily been defined for B clade responses. The C clade peptides that trigger immunodominant responses tend to be localized in the same regions (Johnston *et al.*, 2001). Selecting a clade-appropriate vaccine for regional trials tends to increase the number of potentially cross-reactive epitopes by increasing the level of similarity between the vaccine and the population and the use of consensus and ancestral sequence would enhance cross-reactivity potential (McMichael *et al.*, 2001;Johnston *et al.*, 2001).

In regions where the epidemic is dominated either by a particular subtype or circulating recombinant forms (CRF), use of a dominant lineage for vaccine and use of consensus or ancestor is important. In regions where two or three subtypes and multiple recombinants co-circulate, including each of the prevalent subtypes could improve the potential coverage of not only those subtypes, but of the variety of recombinant forms that stem from them (Angwale *et al.*, 2002). Vaccines specifically designed to target these conserved epitopes if successful, may ultimately optimize by fine-tuning subtype-specific vaccines (Gaschen *et al.*, 2002). If a single subtype predominates in a country, combining an M-group consensus with a regionally dominant subtype might be advantageous in an urban context where people of many nationalities mingle.

2.11.4 Polyvalent Vaccine Approach to Overcome HIV-1 Diversity

Another concept is the use of multivalent cocktails of proteins that include a spectrum of regional variants. Several polyvalent formulations have been designed to contend with HIV-1 diversity. Slobod *et al.* (2003) presented an approach with a 23-valent vaccine (polyEnv1), including variants from several clades, as well as variants with different neutralization resistance profiles and monoclonal antibody binding profiles. Nations like Democratic Republic of Congo with diverse viral population can be best served by developing polyvalent vaccines including a spectrum of natural forms combined with an M-group consensus (Mokili *et al.*, 1999; Vidal *et al.*, 2000)

All these approaches assume that immune responses elicited by one circulating strain will be sufficiently cross-reactive to protect against other strains from the same subtype (Gaschen *et al.*, 2002). Consensus sequences and strains from viable isolates could be combined in a polyvalent approach. Modified envelopes can also be designed to enhance exposure of epitopes known to be capable of inducing broadly neutralizing antibodies (Barnett *et al.*, 2001; Binley *et al.,* 1996; Zwick *et al.*, 2001).

2.12 Prediction of Peptides Binding To Major Histocompatibility Complex Molecule

T-cells are the key components of adaptive immune system, playing a pivotal role in fighting both infectious and cancer cells (Paul, 1998; Walker *et al.*, 1998). T-cell-based immune responses are driven by antigenic peptide (epitope) presented in the context of MHC molecules (Zinkernagel *et al.*, 2004). Therefore, the prediction of peptides that bind to MHC molecules has become the basis for the anticipation of T-cell epitopes (Brander, 2004). Successful vaccination leads to expansion of a set of T-cells that are specific for T cell epitopes contained within the protein sequence of the vaccine and establishment of T-cell memory to those epitopes (Seth *et al.*, 1998).

Recognition of a foreign antigen by T-cells requires that the antigen derived peptides be displayed within the context of an MHC molecule binding pocket (De Groot *et al.*, 2001; Figure 2.4). MHC molecules fall into two major classes, MHC-1 and II. Antigens presented by MHC-1 and II are recognized by two distinct sets of T-cells, CD8+ T-cells and CD4+T-cells, respectively (Brander, 2004). On the surface of the APC, the MHC class1 peptide complex interacts with the T-cell receptor of class 1 restricted T-cell (usually CTLs) and the MHC class II-peptide complex interacts with the T cell receptor of class II restricted T-cell. This triggers a cascade of events that culminate in T-cell response (De Groot *et al.*, 2001)

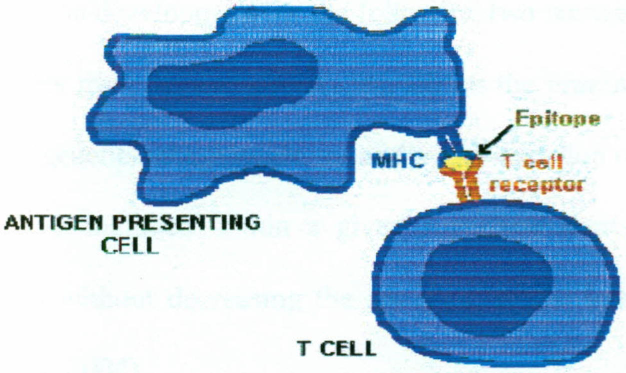


Figure 2.4: The interaction between the MHC, peptide and T-cell (Adapted from De Groot *et al.*, 2001).

In humans MHC molecules are known as Human Leukocyte Antigens (HLA) and there are hundreds of allelic variants of the class-1 (HLA-1) and class II (HLA-II) molecules These HLA allelic variants bind distinct sets of peptides as MHC polymorphism is the basis for peptide-binding specificity (Zinkernagel and Hengartner, 2004) and are expressed at variable frequencies in different ethnic groups.

Human immunodeficiency virus-1 (HIV-1) amino acid sequence polymorphisms associated with expression of specific human histocompartibility leukocyte antigen (HLA) suggest sites of cytotoxic T-lymphocyte (CTL)-mediated selection pressure and immune escape (Brander, 2004). CTLs are HLA restricted, as they only recognize peptide when presented in the context of the appropriate HLA molecule (William *et al.*, 2003).

Groups of HLA molecules (Supertypes) can bind largely overlapping sets of peptides (Settle *et al.,* 1999). The identification of these HLA super types facilitates the epitope based vaccine development for the following two reasons: first targeting of representative HLA alleles from distinct supertypes, allows the immune response to be stimulated in a variety of genetic backgrounds; secondly, the selection of promiscuous peptide binders to these alleles included within a given supertype limits the number of peptides to be considered without decreasing the spectrum of the immune response (Zinkernagel and Hengartner, 2004).

2.13 Designing a Cross-Clade, Epitope Driven HIV-1 Vaccine Using Bioinformatics

The epitope driven vaccine concept is an attractive approach and is successfully pursued in a number of laboratories (Hanke. *et al.*, 1998; An and Whitton, 1997; Nardin *et al.*, 2001). A number of researchers including the groups working in the TB/HIV Research laboratory have been promoting and pursuing the development of a novel HIV-1 vaccine that reflects the global diversity of HIV-1 strains and the HLA variability of human populations (HIV Molecular Immunology Database lanl.1995; Cease *et al.*, 1994; Los Alamos HIV database, 1995; Wilson *et al.*, 2001). A major complication to vaccine development approach is the extreme polymorphism of the MHC molecules (Reche and Reinherz, 2005). Identification of HLA-1 and 11 restricted epitopes (T-cell epitopes) is important for both understanding disease pathogenesis and vaccine design (De Groot *et al.*, 2003). Thus, the availability of computational methods that can readily identify potential epitopes from primary protein sequences has fueled a new paradigm in vaccine development that is driven by this epitope discovery (De Groot *et al.*, 2003).

Designing a cross-clade, epitope driven HIV-1 vaccine using bioinformatics consists of searching for conserved peptide sequence, determining whether they are immunogenic and then combining these epitopes together in a multi epitope vaccine (De Groot *et al.*, 2001). Sophisticated sequence tools to search protein sequence for T-cell epitopes have been available for approximately 10 years (Brander and Goulder, 1999). However Bioinformatics tools to map across-clade B-cell epitopes do not yet exist (De Groot, *et al.*, 2001). The combined use of several bioinformatics tools makes it possible to analyze published variants of HIV-1 genome and identifying both class I and II restricted T-cell epitopes for use in epitope driven HIV-1 cross-clade vaccine.

A typical epitope-based vaccine construct contains a single start codon with coding sequences for epitopes inserted consecutively in the construct, with or without intervening pacer amino acid (Nardin *et al.*, 2001). Selected peptides are synthesized and both MHC binding capability and T-cell responses to the peptide can be evaluated *in vitro* using T2 cell binding assay (Ljunggren *et al.,* 1990). T cell response to the peptides can be measured in standard gamma-interferon release ELISpot assay (Lieberman *et al.*, 1997).

2.14 Tools for Identifying T-Cell Epitopes

Selection of peptides that are highly likely to bind to MHC will enable the identification of both T-helper and cytotoxic T-cell (CTL) epitopes, which are critical components of HIV vaccine research (De Groot, 2001). Utilization of computer algorithm by HIV vaccine researchers may reduce the cost and labour associated with HIV vaccine research by 2-20 fold (De Groot *et al.*, 2003).

There are different computer driven algorithms (web servers) used to predict peptide binding to MHC (De Groot *et al.*, 2003). The discovery of MHC binding motif prompted the development of new bioinformatics tools for vaccine design (Parker *et al.*, 1994; Lipford *et al.*, 1993; Rotzsche and Falk, 1991). This includes: PEPVAC (promiscuous Epitope-based vaccine) that allows the prediction of epitopes to five HLA supertypes, A2, A3, A24, B7 and B15, Epipredict that predicts HLA 1 restricted T-cell epitopes, Vaccinome based on TEPIPOPE algorithm for MHC 11, SYFPEITHI motif matrix for both MHC1 and 11 molecules. , MHCpred, Epimer, Amphi, Epimatrix and Biovation.

CHAPTER THREE: MATERIALS AND METHODS

3.1 Study Population and Site

The study population consisted of volunteers attending VCTs, antenatal clinics attendees, from refugee sites, cases of STDs, blood donors, TB patients and children born of HIV positive mothers. Subjects enrolled were HIV infected individuals recruited from three locations in Northern Kenya that is Moyale, Mandera and Turkana districts as part of on going KEMRI projects in these areas. In Moyale and Mandera patients were recruited from the out patient department of the District Hospital or the inpatient already tested positive for HIV and counselled within the health facility framework. In Turkana district, subjects were selected from known HIV positive patients attending the AMREF supported health facility supported by UNHCR. The samples were analysed at the Centre for Virus Research, KEMRI, Nairobi.

3.3. Sample Size Determination

The sample size was calculated using the formula described by Fisher *et al.,* (1988). The general HIV prevalence according to the 2003 National AIDS Control Council estimate of 9.4% (http://www.kenyaweb.com/health/aids/nationalpolicy.html#nacc) was used. Hence, the sample size required was calculated using the following formula:

$$N = \frac{Z^2{}_{1-\alpha}\, p\,(1-p)}{d^2}$$

Where $Z^2{}_{1-\alpha}$ is the two sided standard error corresponding to 99% confidence interval = 1.96; d^2 is the absolute precision = 5% [0.05 the inverse of 95% confidence limit (the allowable error)] and P is the proportion of prevalence values of HIV in Kenya as by the year 2003 =0.094 (9.4%) as shown by National AIDS Control Council. Substituting for the proportionate values, a total of 139 subjects were obtained.

Therefore:

$$N = \frac{1.96^2 (0.94)(0.06)}{0.05^2} = 139.2369402 = 139 \text{ samples}$$

A total of 139 samples were used.

3.4 Ethical Considerations

Clearance to carry out the study was obtained from Kenyatta University. Approval for the study was obtained from the scientific steering committee and ethical review committee of KEMRI.

3.5 Sample Collection

Approximately 5ml of venous blood was collected in heparine containing tube and transported in a cooler box containing ice. Once in the lab, the samples were stored at -20 before processing for Peripheral Blood Mononuclear Cells (PBMC) extraction.

3.6 Deoxyribonucleic Acid (DNA) Extraction

3.6.1 Extraction of Peripheral Blood Mononuclear Cells (PBMC) from Whole Blood

Peripheral Blood Mononuclear Cells were isolated from 5ml of whole blood according to the method described by Chomczynski *et al.* (1997). Briefly, falcon tube containing 10 ml of 0.84% ammonium chloride, 10 ml of 0.84% ammonium chloride was added to 3-5 ml of whole blood and vortexed to mix completely. This was then incubated at 37°C for 10 minutes and spinned in a centrifuge at 350g for 10 minutes and supernatant discarded. 10ml of 0.84% ammonium chloride was added to the resulting pellet and the procedure above repeated again till the pellet appeared clear. 1ml of 0.84 ammonium chloride was added to the pellet then was drawn into an Eppendorf tube and centrifuged for 10 minutes. The supernatant was sucked off leaving white PBMC pellet at the bottom, which was dried and stored at -20°C until use.

3.6.2 Proviral HIV Deoxyribonucleic Acid Extraction from Peripheral Blood Mononuclear Cells

Proviral HIV DNA was extracted from PBMCs according to the method described by Chomczynski *et al.* (1997). Briefly 500µl of DNAzol genomic DNA extraction reagent (Gibco BRL®) was added to dissolve the DNA. Two volumes (1000µl) of chilled absolute ethanol (cooled to 4°C) was added to precipitate the DNA and centrifuged at 300g for 10 minutes. The precipitated DNA was washed twice with 1ml of 70% ethanol, the supernatant discarded and the pellet dried. The DNA was then solubilised in 100µl of distilled DNAse /RNAse free water.

3.7 Polymerase Chain Reaction (PCR) Amplification Using Proviral DNA

The starting template for amplification was proviral DNA that had been extracted from whole blood as described above. Polymerase Chain Reaction (PCR) was carried out using methods described by Carr *et al.* (1999). The primers that were used in the study were specifically for the *gp 41 env*, *gp 120* (C2V3) and the *P24 gag* region of HIV-1 (Table 3; Heyndrickx *et al.*, 2000). A master mix containing 2mm MgCl, 2mmdNTPS, 0.5 units Taq polymerase, 1× PCR Buffer, 2ng of each primer (forward and reverse) and DNA template was made. Constitution of the master mix was performed on ice to minimize the commencement of any reactions. Taq polymerase was added last as it is sensitive to the buffer and also reaction begins immediately it is added. Two consecutive PCR assays were performed making the PCR highly specific. Two sets of primers were used for the 1st and 2nd round PCR respectively. Table 3.1 shows the sequences of the primers used in the study and the region of amplification. The PCR thermocycling conditions using these primers were: Hot start at 95°C for 10 minutes, denaturation at 95 °C for 30 seconds, followed by annealing at 55 °C for 30 seconds, and extension at 72 °C for 60 sec, with a final extension at 72 °C for 7 minutes. A total of 35 cycles were carried out.

3.8 Agarose Gel Electrophoresis of PCR Products

All the PCR products were analyzed by conventional agarose gel electrophoresis as described by Sambrook *et al.* (1989). An agarose gel was prepared on which the PCR products were loaded for visualization of the amplified DNA. The agarose gel was prepared by taking 1 gm of agarose and dissolving in 1x TBE (Tris Borate EDTA) to make a 1 % gel. 5µl of sample was mixed with the 3µl loading dye (Bromophenol blue) and loaded onto the gel, together with the molecular weight marker. Electrophoresis was carried out at a constant voltage of 100volts/cm using Bio-Red model 200/2 power supply source. The gel was then stained with 0.5µl/ml ethidium bromide for 15 minutes and visualized under ultraviolet light. The exact location of nucleic acid fragments on the gel was determined by direct examination of the gel and the size estimated by comparing with the molecular weight marker loaded along side them. The different amplified regions of the DNA strands (*gp 41, gag and gp120*) migrated at different rates according to molecular sizes. The molecular weights of *gp41*, *gp120* and *gag* DNA amplicons products were approximately 400 bp, 600 bp, and 300 bp respectively. This tallied with the expected molecular weight of the expected products (Appendix I).

3.9 Deoxyribonucleic Acid (DNA) Sequencing

Here the starting template was PCR product. Deoxyribonucleic Acid (DNA) sequencing was carried out on all the samples to determine the order of the bases in the PCR products that were amplified directly as described by The Applied BioSystems, USA sequencing manual using BigDye® kit.. The fluorescently labeled dyes were attached to ACGT extension products in DNA sequencing reactions. The dyes come in four colors that is red (labels Thymine bases), blue (labels Cytosine bases), black (labels Guanidine bases) and green (labels Adenine bases). AmpliTaq® Polymerase was used for primer extension. To each 3μl of DNA sample, 3μl of 5x buffer and 2μl of BigDye and 1.5μl of primer was added. The sequencing PCR was carried out under the following conditions: 96°C for 5 minutes followed by 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes. After the sequencing PCR, the labeled amplicons were then transferred to sequencing tubes and loaded onto the automated sequencer. During sequencing, a portion of the sample enters the capillary as current flows from the cathode to the anode. When the fluorescently labeled DNA nucleotides reach a detector window in the capillary coating, a laser excites the fluorescent dye labels. A CCD camera collects emitted fluorescence from the dyes. The DNA Sequencing Analysis Software and the GeneScan® Analysis Software programs analyzed the raw data, converting it to DNA sequence.

Table 3.1: Sequences of primers used in the study and the regions of amplification

3.10 Subtype Determination

Region	Primer Name	Sequence	Region on HXB2	Reference
Env *gp* 41	*gp* 40F1: forward 1	TCTTAGGAGCAGCAGGAAGCACTATGGG	7789–7816	Carr *et al.*, 1998
	gp 41R1: reverse 1	AACGACAAAGGTGAGTATCCCTGCCTAA	8347–8374	Carr *et al.*, 1998
	gp 46F2: forward 2	ACAATTATTGTCTGGTATAGTGCAACAGCA	7850–7879	Carr *et al.*, 1998
	gp 47R2: reverse 2	TTAAACCTATCAAGCCTCCTACTATCATTA	8281–8310	Carr *et al.*, 1998
Gag p24	*gag* F1: forward 1	TCACCTAGAACTTTGAATGCATGGG	695-715	Heyndrickx *et al.*, 2000
	gag R1: reverse 1	CTAATACTGTATCATCTGCTCCTGT	1786-1810	Heyndrickx *et al.*, 2000
	gag F2: forward 2	AAAGATGGATAATCCTGGGTCC	808-830	Heyndrickx *et al.*, 2000
	gag R2: reverse 2	ACATTTCCAACAGCCTTTTT	1475-1495	Heyndrickx *et al.*, 2000
C2V3	M5: forward 1	CCAATTCCCATACATTATTGTGCCCCAGCTGG	6451-6482	Takehisa *et al*, 1998
	M10: reverse 1	CCAATTGTCCCTCATATCTCCTCCTCCAGG	7225-7254	Takehisa *et al*, 1998
	M3: forward 2	GTCAGCACAGTACAATGACACATGG	6541-6566	Takehisa *et al*, 1998
	M8: reverse 2	TCCTTCCATGGGAGGGGCATACATTGC	7114-7140	Takehisa *et al*, 1998

Samples were subjected to PCR using specific primers for the three regions (*gag, env gp 41* and *env gp120 (C2V3*) and then direct sequencing performed. For subtype determination, *gp 41* primers were used. The subtypes of the generated sequences were determined using BLAST reference sequences present on the HIV sequence database http://hiv-web.lanl.gov/content/hiv-db/BASIC_BLAST/basic_blast.html, CLUSTALW and TREEVIEW software.

Sequences from *gp 41* regions were used, this is because it is one of the most variable region and undergoes constant genetic variability (has major antigenic surface epitope variability) due to immunological pressures by the host. The sequences generated were successfully analysed phylogenetically to determine their evolutionary relationships/distance of the nucleotides using CLUSTAL W (version 1.81) and phylogenetic trees constructed using Neibour-Joining method (Saitoi and Nei, 1987) and using references from other regions such as Kenya, Uganda, Tanzania, USA, South Africa and Botswana. Their reliability was estimated by 1000 bootstrap replications. The closest subtypes had a higher bootstrap value as compared to the one that were not closely related. The profile of the trees was visualised with Tree view PPC version 1.65 (Institute of biomedical and life science, Scotland, UK).

3.11 Limitations

Some samples did not amplify using the primers used. For such cases, amplification was repeated using a different primer for another region. Some of the sequences generated did not cluster with any reference sequences. For these sequences, their subtypes were determined by direct BLAST analysis (www.hiv.lanl.gov/content/hiv_db/BASIC_BLAST/basic_blast.html).

3.12 Phylogenetic Analysis of Generated HIV-1 Sequences

Sequences generated in this study were analyzed phylogenetically using the BLAST software (Altschul *et al.*, 1997) to determine their subtype and how closely related sequences were (Karlin *et al.*, 1990; Karlin and Altschul, 1993). Phylogenetic constructions were generated by using Neighbour-Joining (NJ) method (Siatou and Nei, 1987) and CLUSTAL W program (Thompson *et al.*, 1994). The trees were visualized using the TREEVIEW program (Page, 1996).

3.1.4 Identification of Conserved Amino Acid Epitopes

All the 139 sequences generated were translated into amino acids using Translation for Publication program and successfully aligned according to the different subtypes to which they belonged using the CLUSTAW program (Version 1.81) to determine their level of conservation. The amino acid sequences were then used to identify conserved amino acid regions in the generated sequences. The sequences from the three regions (*gp* 120, *gp* 41 and *gag*) were aligned according to the different subtypes to which they belonged (Figures 4.6, 4.7 and 4.8; Appendix II). This was done to determine if there were any amino acid differences between the subtypes and the conserved regions were identified.

3.15 Epitope immunogenicity Prediction Using SYFPEITHI

Once the conserved epitopes were identified, a BLAST analysis of the conserved sequences onto SYFPEITHI (a computer aided vaccine Design program that predicts immunogenicity at the level of epitope and prediction of peptide binding affinity to MHC for identification of T-cell epitopes) was carried out (Rammensee *et al.*, 1999; www.syfpeithi.de). This analysis was done for the different subtypes by selecting the conserved epitopes and analyzing them using the SYFPEITHI program. This epitope prediction tool is able to predict epitopes recognized by CD8+ CTLs that are immunogenic and also show their HLA restriction.

SYFPEITHI yields a score for each peptide and is scored in a 9-mer frame. Scoring is a quantitative estimate of the likelihood (relative to the sequence) that a peptide will bind to a given HLA molecule. The scoring system evaluates every amino acid within a given peptide. Individual amino acids may be given the arbitrary value1 for amino acids that are only slightly preferred in the respective position, optimal anchor position are given value15 and any value between the two is possible (Rammensee *et al.*, 1999).Negative values are also possible for amino acids, which are disadvantageous for the peptide binding capacity at a certain sequence position. SYFPEITHI also indicates silent regions of epitopes that elicit poor or no immune response. The epitopes that are not immunogenic are given a score of up to negative one (-1) and the best epitope is expected to have a score of 36 (the highest score that can be predicted by the SYFPEITHI program (Rammensee *et al.*, 1999). The maximal score varies between different MHC alleles for example the maximal score for HLA-A*201 peptides is 36 (Rammensee *et al.*, 1999). For this study epitopes that had a score of less than 10 were considered poorly immunogenic while the epitopes with a score of greater or equal to 20 were considered immunogenic. Using this epitope prediction tool, the regions of an individual epitope that are most immunogenic are indicated by in bold (bolded) and the silent regions that do not elicite a strong immune response is not bolded. While the regions that are underlined are more immunogenic but not as immunogenic as the ones bolded.

3.16 Design of Cross-Clade Multiepitope Candidate Vaccine

In this study, epitopes with highly significant scores (above 25) using SYFPEITHI were selected and used to develop a candidate DNA vaccine that is 300 amino acid long. These conserved immunogenic epitopes were identified from subtype A1, D and C in gag, gp 120 and gp41 regions of HIV. The epitopes were put together in order to construct a multi-epitope (super-epitope) candidate vaccine is cross-reactive to protect against the major clades circulating in Kenya.

3.16 Data Analysis and Presentation

Nucleotide sequences generated in this study were analysed phylogenetically and subtypes are presented in form of pie charts. The sequences generated were translated into amino acid using Translation for Publication program and aligned to determine their level of conservation using ClustalW (version 1.81). The identified conserved regions were analysed using SYFPEITHI –an epitope prediction software tool (Rammensee *et al.*, 1999) to establish their immunogenicity.

CHAPTER FOUR: RESULTS

4.1 Subtype Determination

For subtype determination, 86 samples were successfully sequenced using the *env gp 41* primers. The three border districts namely, Turkana, Moyale and Mandera districts were used to determine the subtypes of HIV circulating in Northern Kenya (Figures 4.1, 4.2 and 4.3).

From the 86 *gp 41* sequences, a total of five phylogenetic trees were generated from the sequences (Figures 4.4–4.8). The sequences were analysed in groups of between 9 and 20 to avoid overcrowding when constructing Phylogenetic trees. Some of the sequences did not cluster with any reference sequences such as MYDH016, MADH062 and MADH 010. For these sequences, their subtypes were determined by direct BLAST analysis (www.hiv.lanl.gov/content/hiv_db/BASIC_BLAST/basic_blast.html).

In this study, analysis of the samples from the three Districts showed that 44% of the sequences generated from the three districts were subtype A1, 45% were subtype C and 11% were subtype D.

In Moyale district a total of 56 samples were successfully sequenced using *env gp41* primers. 36% of the samples were subtype A1, 55% were subtype C and 9% were subtype D (Figure 4.1). Therefore HIV-1 subtype C was the most dominant subtype in Moyale.

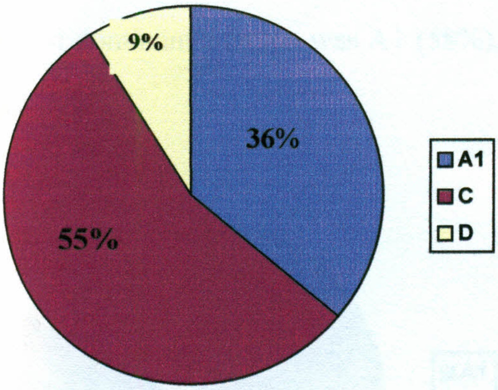


Figure 4.1: The Prevalence of HIV-1 subtypes in Moyale
Showing the distribution of HIV-1 subtypes A1, C and D in Moyale

A total of 6 samples from Mandera were successfully sequenced using *env gp 41* primers. Phylogenetic analysis showed that 67% were subtype A1, 33% were subtype C (Figure 4.2). There was no subtype D in this region. HIV-1 subtype A1 was the most dominant subtype in Mandera.

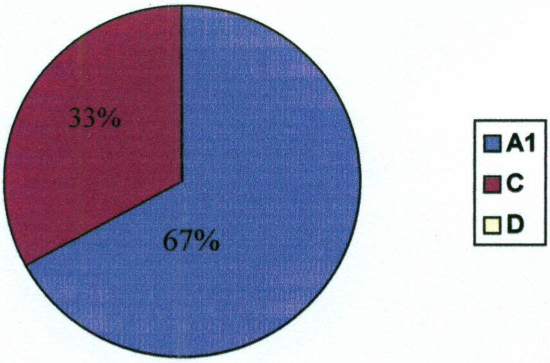


Figure 4.2: The Prevalence of HIV-1 subtypes in Mandera
Showing the distribution of HIV-1 subtypes A1, C and D in Mandera

In Turkana a total of 24 samples were successfully sequenced using *env gp 41* primers. The most dominant subtype was A1 (58%), subtype C was 25% and subtype D was 17% (Figure 4.3).

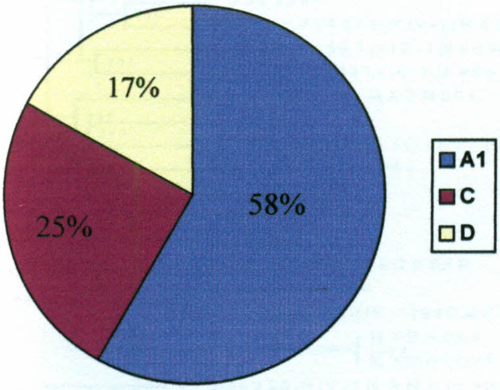


Figure 4.3: The Prevalence of HIV-1 subtypes in Turkana
Showing the distribution of HIV-1 subtypes A1, C and D in Turkana

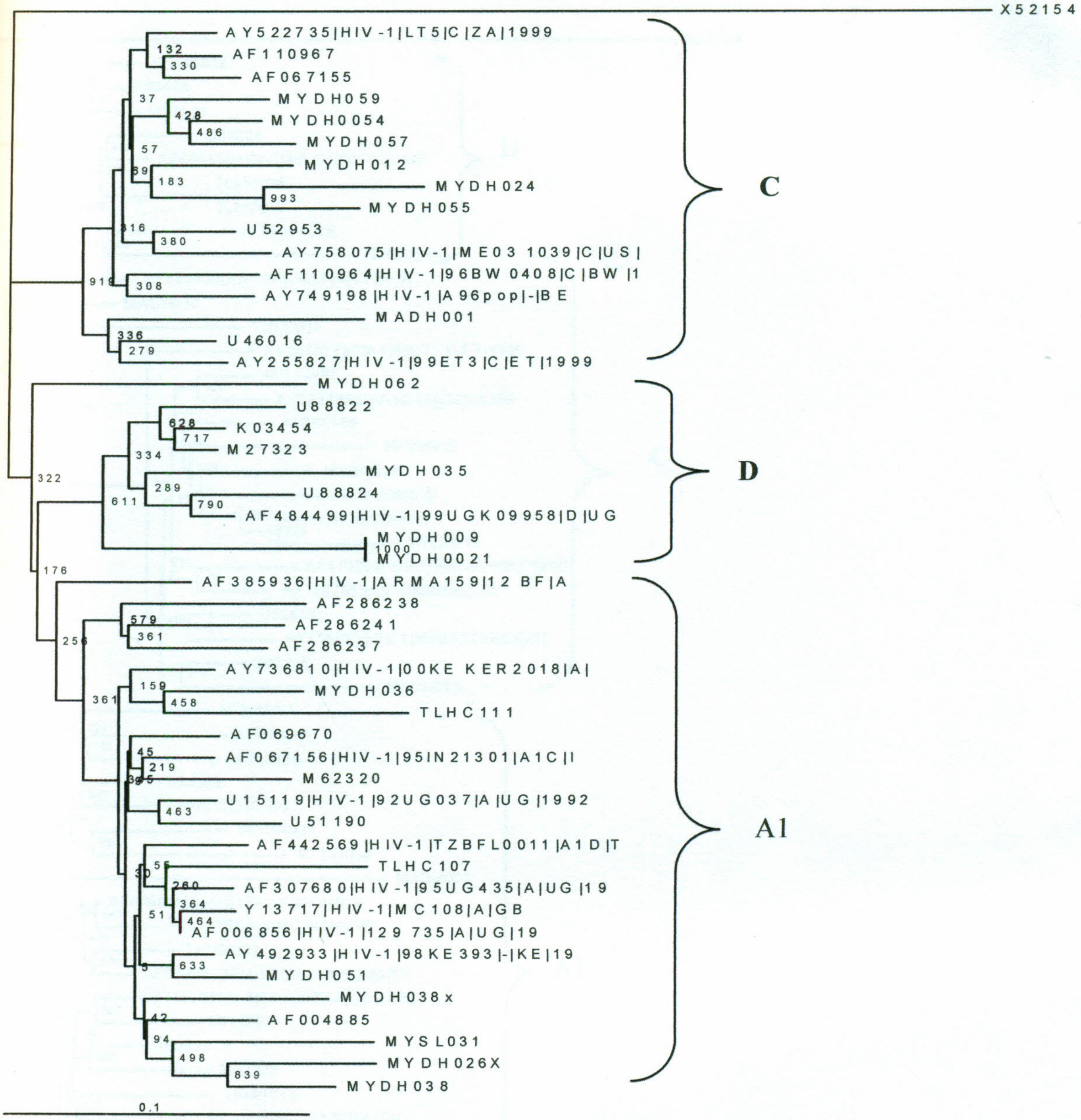


Figure 4.4: Phylogenetic tree representing sequences 1-19.
The sequences in this tree clustered with HIV-1 subtype A1, C and D reference sequences from Tanzania Uganda, Gabon, Ethiopia, South Africa, Botswana and The United States.

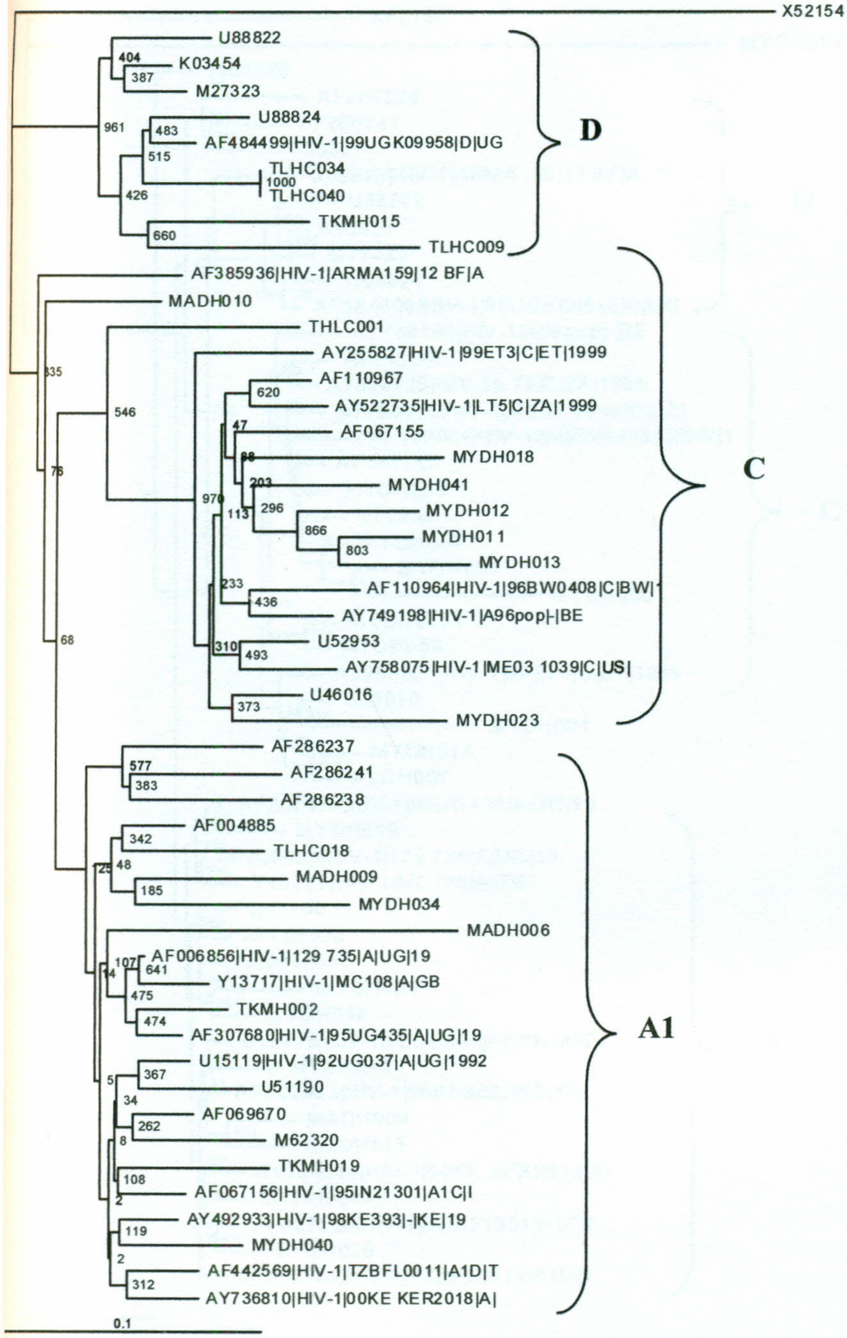


Figure 4.5: Phylogenetic tree representing sequences 20-38. The sequences in this tree clustered with HIV-1 subtype A1, C and D reference sequences from Tanzania Uganda, Gabon, Ethiopia, South Africa, Botswana and Belgium

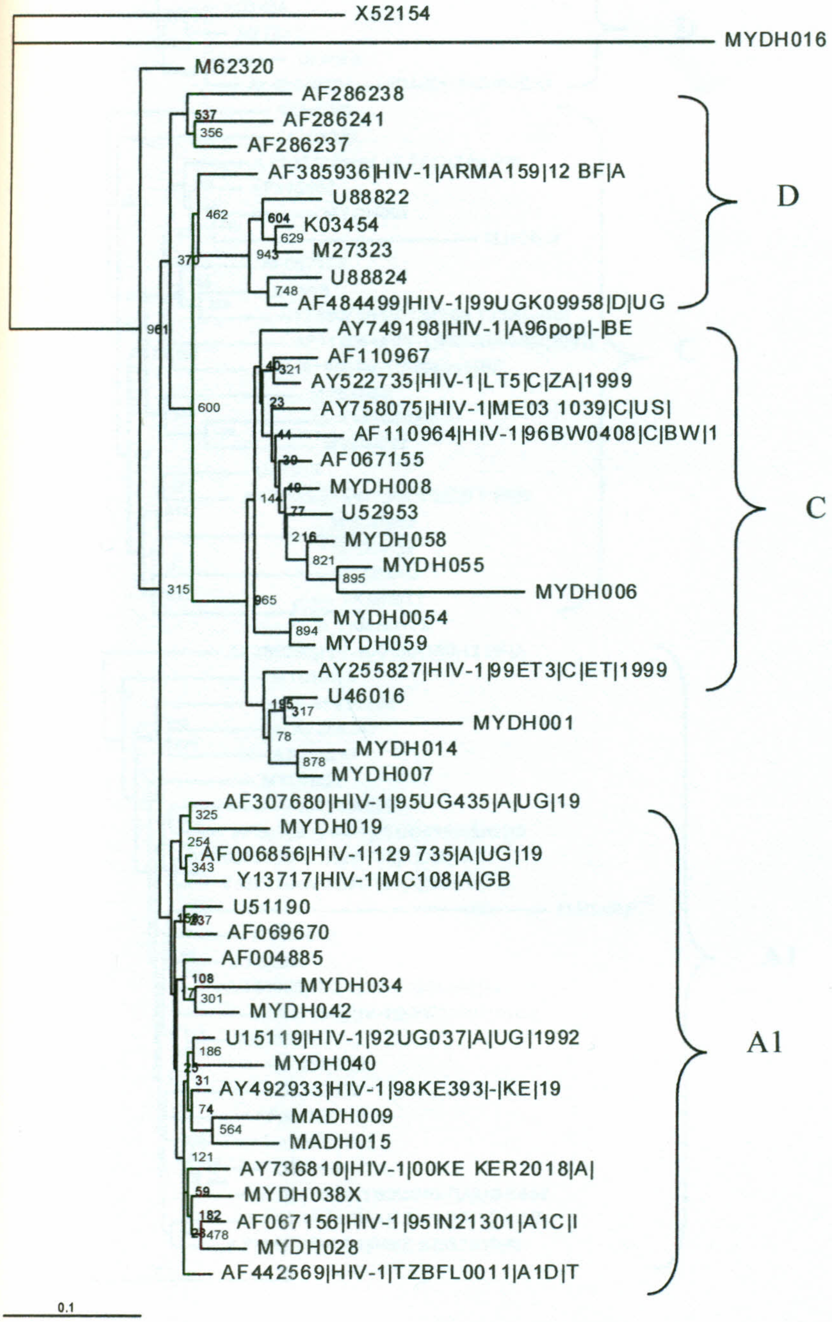


Figure 4.6: Phylogenetic tree representing sequences 39-56
The sequences in this tree clustered with HIV-1 subtype A1, C and D reference sequences from Tanzania ,Kenya, Belgium, Uganda, Gabon, Ethiopia, South Africa, Botswana and The United States.

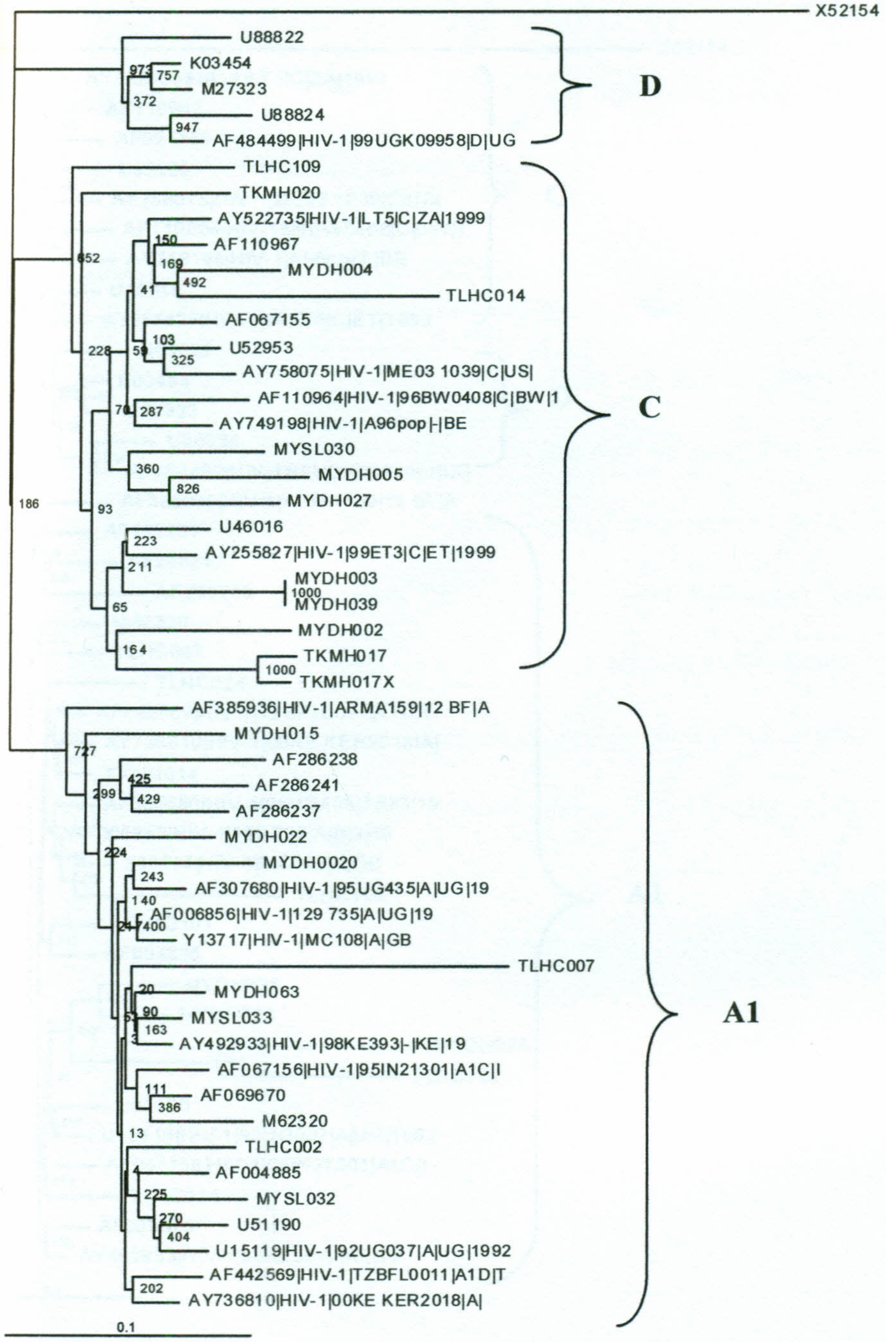


Figure 4.7: Phylogenetic tree representing sequences 57-76. The sequences in this phylogenetic tree clustered with HIV-1 subtype A1, C and D reference sequences from Tanzania ,Kenya, Belgium, Uganda, Gabon, Ethiopia, South Africa, Botswana and The United States.

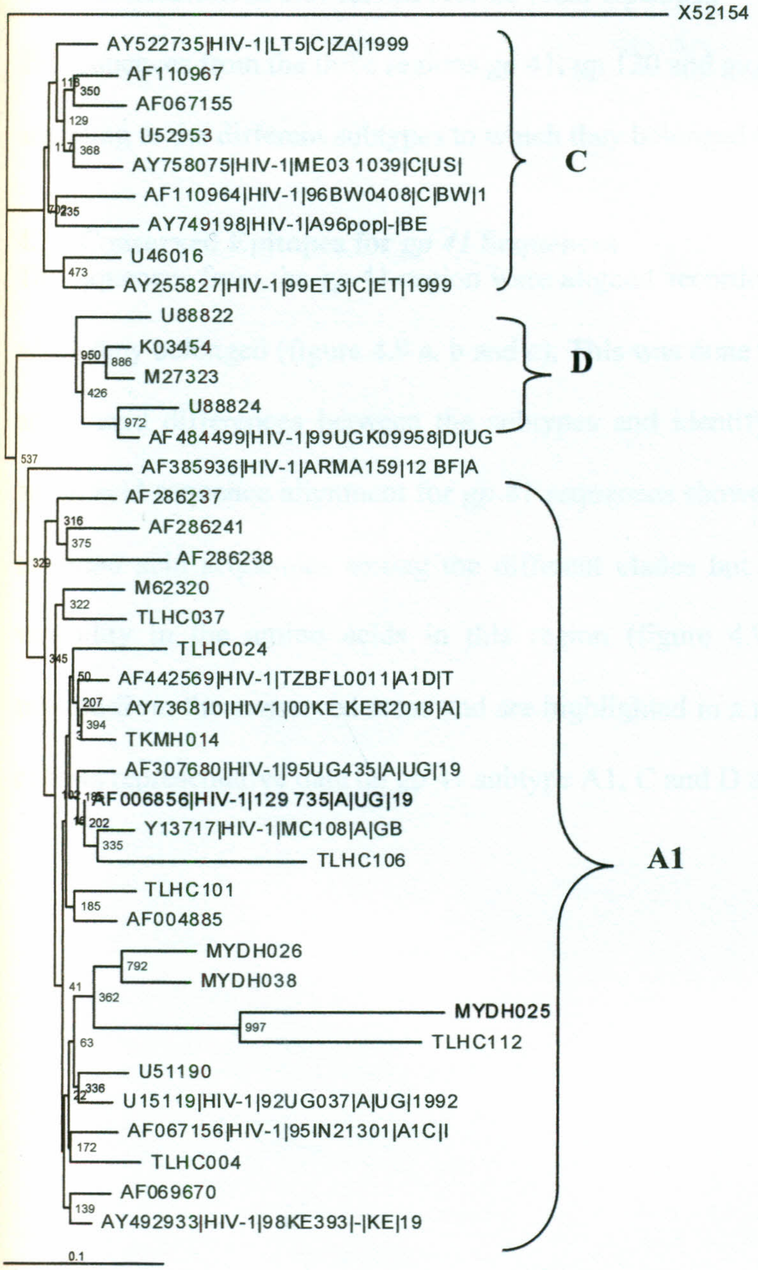


Figure 4.8: Phylogenetic tree representing sequences 77-86. The sequences in this tree clustered with HIV-1 subtype A1, C and D reference sequences from Tanzania, Kenya, Belgium, Uganda, Gabon, Ethiopia, South Africa, Botswana and The United States.

4.2 Identification of Conserved Amino Acid Epitopes

The sequences from the three regions *gp* 41, gp 120 and gag regions were aligned according to the different subtypes to which they belonged (Appendix II).

4.2.1 Conserved Epitopes for *gp 41* Sequences

The sequences from the *gp* 41 region were aligned according to the different subtypes to which they belonged (figure 4.9 a, b and c). This was done to determine if there were any amino acid differences between the subtypes and identify the conserved regions. The amino acid sequence alignment for *gp* 41 sequences showed that there was conservation of amino acid sequences among the different clades but some could not align due to variability in the amino acids in this region (figure 4.9). Conserved epitopes were selected from the conserved areas and are highlighted in a red colour. Figure 4.9 a, b and c shows representative data on gp 41 subtype A1, C and D alignment.

a) gp 41 subtype A1 alignment

```
MYDH063_        ------FEAIEA-QQHLLKLT-----VWGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 47
TLHC024_        -GGLTFAEGYRG-STASVETH-----GLGHSTAPGKILALER-YLKDQQLLGIWGCSGKL 52
MYSL030_        ----------KA-QQHMLQLT-----GLGHKQLQTRVLALER-YLKDQQLLGIWGCSGKL 43
MADH009_        -GAYNLWK-AQRLNNICCKLT-----VWGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 52
MYDH022_        -GAYNLWK-AQRLNNICCKLT-----VWGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 52
MYDH040_        ------LR-LSRLNNICCKLT-----VWGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 47
TLHC004_        ----NLFGGLSRLNNICSNSR-----SRGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 50
MYDH019_        -----LLKVLEA-QQQLLRLT-----VWGIKQLQARVLALES-YLRDQQLLGIWGCSGRL 48
MYDH015_        ---NFVGRLLEA-QQHMLRSR-----SRGIKQLQARVLALER-YLRDQQLLGIWGCSGKL 50
TKMH014_        -------EGIEA-QQHLLKLT-----VWGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 46
TLHC002_        LGDQNLWRAIEA-QQQMLRLT-----VWGIKQLQARVLAVER-YLKDQQLLGIWGCSGRL 53
MYSL033_        -WGHNLLRAIEG-STTSVETH-----SLGHKQLQARVLAVER-YLKDQQLLGIWGCSGKL 52
MYDH0020_       LEHYNFAGGLQRLNNICYKLT-----VWGIKQLQARVLAVER-YLKDQQLLGIWGCSGKL 54
MYSL032_        ----ICWRAIEA-QQHLLKLT-----VQGHQTAPGKSPRLWKDTLRDQQLLGIWGCSGKL 50
MYDH034_        ---------IEG-STTSVETY-----CLGHYNSSRQESWLWKDTSRINNSSGIWGCSGKL 45
TLHC101_        VFGGIEFVRLRG-STTSVETH-----GLGHSNSSRHCVLAVERYLKDQQLLGIWGCSGKL 54
MADH015_        --GDIIFGGLSRLNNICCNSR------SGALNSSRHSVLAIERYLRDQQLLGIWGCSGKL 52
MYDH042_        -------CGYRRLNNICRSSR------SGALNSSRQESWLSKDTSEDQQLLGIWGCSGKL 47
MYDH025_        ------LCGLYMLNNISLQTHGSRGHYTAPCKIPRAGGKDTWMGSTAPRNLGAGSGKTHL 54
TLHC037_        ---------LYRLNNICC-TH-SLGHYTAPGKSPGCG--KIPKGSTAPRNLGL-LWKTHL 46
MYDH026X_       -LLGPAIFGGCSSTTSVETHG------LGALNSSGQESLLWK-DTYGINSSSEFGAALES 52
MADH010_        -----NSVGLRGSTAYVETHS------LGHKQLQAR-SWLWK-DTERINSSYEFGAALED 47
MYDH038_        -------SRLSRLNNICCNSQ------SGALNSSRQESWLWK-DTYRINSSSEFGAALEN 46
MYDH026_        -----NLEGIEAQQHLLKLTV------WGIKQLRARVLAVER-YLRDQQLLGIWGCSG-N 47
MYDH051_        LFGARILXGLELTTSVAIHSL-------GHKQLQARVLGSWKDTSRDQQLLGIWGCSGKL 53
MYDH028_        ----------VEAQQHLLKLT-----VWGIKQLQARVLAVER-HLKDQQLLGIWGCSGKL 44
TLHC106_        --GHTICCRYRGSTAMVGDSR-----SGGIKQLQARVLALER-TYGINSSSVFGGCSGKL 52
MADH006_        ------LTIYAQHHLVKLTCR------ALNISMAMMPMLWEMIPKWSTAPMEFRGTLGKL 48
TLHC007_        ------LTIYAQHHLVKLTCR------ALNISMAMMPMLWEMIPKWSTAPMEFRGTLGKL 48
MYDH024_        ----FWGQRFGGSRQHMLNHG------LGHKQLQTRVLGYRK-IPKGSTAPRDLGLLWKT 49
TLHC112_        -----ILRALEAQQQLLNSRS------GGIKQLQARVLAVGKIPKGSTAPRNLGAALENS 49
```

b) *Gp* 41 Subtype C Alignment

```
MYDH0054_       ----LRVYEA-QQLIVETHSLGALSSSRQESSALERYLKDQQLLGIWGCSGNSSAPLLCL 55
MYDH059_        ----LEAIRG-ATAYVATHSLGHLSSSRQESWLSKDTYRINSSSGFGAALENSSAPLLCL 55
TLHC109_        PSIFAKGYRKRNSRMFATHGLGASNSSRRESWLWKDTSGINSSSGFGGCSGKLICTTNVP 60
MYDH014_        ----NLLRLLEAQQHMLKLTVWGIKQLQTRVLAIERYLQDQQLLGIWGCSGKLICTTAVP 56
MYDH007_        ----NLLRAIEAHSIQLQLTVWGIKQLQTRVLAIERYLQDQQLLGIWGCSGKLICTTAVP 56
MYDH004_        ----ICLRAIEAQQHLLQLTVWGIKQLQTRVLAIERYLKDQQFLGIWGCSGKLICTTSVP 56
MYDH008_        -------EGIEAQQHMLQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVP 53
MYDH018_        --------GIEAQQHVLQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVP 52
MYDH023_        ------GDYRRRNSIYSQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVP 54
MYDH055_        ------EGYRGATAIVATHGLGHPKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVP 54
MYDH039_        ------EGYICATSYVATHGLGH-YSLHARFLAIEIYLKDQQLLGIWGCSGKLICTTAVP 53
MYDH003_        ---NLFGGYRGASAYVATHGLGHLSSSRQESQAIERYLQDQQILGIWGCSGKLICTTNVP 57
MADH001_        --------PRGATAYSSNSHSRGIKQLQTRVLAIERYLKDQQLLGIWGCSGNSSAPLMCL 52
MYDH002_        --------PRGATAYSSNSHSRGIKQLQTRVL
AIERYLKDQQLLGIWGCSGNSSAPLMCL 52
MYDH001_        -----ICCGIEAQQHMLQLTVYGIKQLQARVLALERYLRDQQLLGMYGCSRNSSAPLLCL 55
TKMH020_        ------WKAIEAQQHMLQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGNSSAPLMCL 54
MADH013_        -------------HIVATHSLGALSSSRQESLALERYLKDQQLLGIWGCSGKVVCTTAVP 47
MYDH005_        -------------HIVATHSLGALSSSRQESLALERYLKDQQLLGIWGCSGKVVCTTAVP 47
MYDH041_        ------LKVSKANNICCNSRFRGIKQLQTRVLALERYLKDQQLLGIWGCSGKLICTTAVP 54
TLHC014_        ---------FQRRNSVVATHSLGHKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVP 51
TKMH017_        --------TIEAHSICCSSRSGALSSSRRESWLLEKYLKDQQLLGIWGCSGKLICTTTVP 52
TKMH017X_       -------ETIERNSICCSSRSGALSNSRRESWAIEKYLKDQQLLGIWGCSGKLICTTTVP 53
MYDH006_        --SDFAKGLKGATALFASHGLRHLSSSRQAVLAIKKYLKNQPLLGISGCSCKTHLHHCCT 58
MYDH012_        -------GYRGATAMLQLTVW-GIKQLQDKSPSYRKIPKGSTAPRTLGLLWKTHLHHCCA 52
MYDH011_        ------EGYRGATAYSCNSRS-GALSSSRQESSLWKDTQRINSSSEFGAALENSSAPLLC 53
MYDH013_        ------EGYRGATAYSCNSRSRGIKQLQTRVLALGKIPKGSTAPRNLGLLWKTHLHHCCA 54
MYDH058_        -----FEAIEGATAYVA-THGLGHSAAPDKSPGYRKIPKGSTAPRDLGLLWKTHLHHCCA 54
MYDH027_        TGATICWTLRGATALLA-THGLGIKQLHARVLALERYLKDQQLLRDLGLLWKTHLHHCCA 59
                                                               .   :
```

c) *Gp* 41 Subtype D Alignment

```
MYDH057_        ------FAEGITRPTAYVALTVRGIKQLQTRVLAIER-YLKDQQLLGIWGCSGKLICTTA 53
TLHC034_        ------FAEGYRRRNTSVATHCLGHKQLQARVLAVES-YLRDQQLLGVWGCSGKHICTTA 53
MYDH062_        ---------RYRGATAYVRTHSLGHYTAPGKSPGCRKGYLKDQQLLGIWGCSGKLYLQPL 51
MYDH035_        -----ICCGAIRRRTSSVATHSLGRYTAPGKNPGCGK-NLKDQQLLGIWGCSGSHICTTT 54
TLHC009_        -----TFHEGYQRRSSFVATHSLGLSNSSKQESWLWK-GPSRSTAPRNLGFSGSTLFRTT 54
TKMH015_        -----NICLRAIEAQQLCCNSQSGALNSSRQESWLWK-TTQGPTAPRNLGFSGRHICTTN 54
MYDH0021_       FGASEFFCGYQRLNNICWKLTVWGIKQLQGTESWLWK-DTSGINSSSEFGVALENIFAPL 59
TLHC040_        -----IFCDYRGAT-ALLQLTVWGIKQLQ-ARSWLWK-ATYRINGSSEFGVALENTFAPL 52
MYDH009_        -----NIAEGIEAQQHCCNSQSGALNSSR-RESWLWK-DTYGINSSSGFGVALEDTFAPL 53
                                       .

MYDH057_        -VPWNSSWSNKSY-QDIWENMTWMQWDREINNYTNTIYRLLGRFAKPAGRK-------CK 104
TLHC034_        -VPWNASWSNKSI-EEIWNNMTWMEWEREIENYTGVIYSLIEESQTQQEKNEQELLQLAN 111
MYDH062_        NVPLELLAGVTSLKKEIWNNMTWMQWEREIGNYTDTIYGLLEESQTQQEMN--------E 103
MYDH035_        -VPWNSTWSNRSL-GEIWNNMTWMQWEKEIDNYTGIIYSLIEASKPSRKRMNKNFWHWTN 112
TLHC009_        NVPWNSTWGNKSL-DKIWN-MTWIQWEKEIDNFPGIISSLIEASKPSRKGMTRNFWNWPM 112
TKMH015_        -VPWTIVGVINHLLEEIWGNMTWMEWEKEISSYTGVIYNLIEESQTQQEKMKKNYWNWTN 113
MYDH0021_       TCPGTLVG-VIKLMVRFGITCPGWSGKEKLTITQVYYIPYLKNRKPSKKKCTRTIGIGPM 118
TLHC040_        LCPGTLAG-VIDLRCHLARTSHGWSGKGRLTITQNSYTASLKNRKSSKKRMKKIYCNWTH 111
MYDH009_        MCHGTLAG-VIKLSSRIWDNMTWMEWEREIDNYTGLIYNLLEKSQAQQEEMNKTYCNWHM 112
                                :       . . .:          :

MYDH057_        GFTLHWDSWDNLWSWLYITNWLWEIKGYS---- 133
TLHC034_        GQVCGIGSHNKMAVDIKIFIMLQELCGFQ---- 140
MYDH062_        QELLPLDKWPNLWNWFDISHWLWIKNFIMKWRL 136
MYDH035_        GQACGIGLHTKVLWIYEYSYCCRRLWG------ 139
TLHC009_        GKLVELVSHTHGCGYKILSSCRSLG-------- 137
TKMH015_        GQVCGMVCHNKLAVDIKIFIMMSELWG------ 140
MYDH0021_       GKFVELVSITNGCGYKNIHNDSRRLG------- 144
TLHC040_        GQVCGLVSHNKCCGYKNIHNDVGGFG------- 137
MYDH009_        GKFVELVNITNWLWISKYSNDSRRLG------- 138
```

Figure 4.9: Conserved Epitopes for *gp 41* Sequences. Showing *gp 41* subtypes A1, C and D amino acid sequence alignment using ClustalW. Conserved epitopes highlighted in red colour

4.2.2 Identification of Conserved Immunogenic Epitopes for gp 41 Subtype A1

Once the conserved sequences were identified, epitopes that were nine amino acids long were picked from the conserved areas. Analysis of the conserved epitopes was done by blasting them onto SYFPEITHI program. Epitopes from the different subtype alignment (Subtype A1, C and D) were analysed separately (Table 4.1, 4.2, 4.3).

A total of 38 *gp 41* Subtype A1 epitopes were identified and analysed using SYFPEITHI program. Out of the 38 epitopes, 17 epitopes with a score ≥20 and the best epitope had a score of 27 (Table 4.1). The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

Table 4.1: Conserved Immunogenic Epitopes for *gp 41*Subtype A1
The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

EPITOPE	HLA RESTRICTION	SCORE
VW**G**I**K**QLQ**A**	HLA-B*08	11
SRGIKQLQA	HLA-B*1402	16
Q**A**RVLAVE**R**	HLA-A*6801	19
G**I**KQLQAR**V**	HLA-A*0201	21*
Y**L**RDQQLL**G**	HLA-A*03	17
G**I**WGCSGK**L**	HLA-A*0201	24*
LSRLNNIC**S**	HLA-A*1101	14
LTIYAQHHL	HLA-A*26	20*
H**H**LVKLTC**R**	HLA-A*6801	17
A**L**NISMAM**M**	HLA-A*0201	19
A**M**MPMLWE**M**	HLA-A*0201	23*
I**P**KWSTAP**M**	HLA-B*0702	20*
EFRGTLGKL	HLA-A*26	24*
YL**K**D**Q**QLL**G**	HLA-B*08	16
Y**L**RDQQLL**G**	HLA-A*03	17
G**I**WGCSGR**L**	HLA-A*0201	22*
N**V**PWNSSW**S**	HLA-A*03	14
SSAPLMCP**G**	HLA-A*1101	15
M**C**PGTLVG**V**	HLA-A*0201	19
GVINLKRRY	HLA-A*26	25*
K**R**NLAITQ**I**	HLA-B*2705	25*
K**E**IDNYTQ**L**	HLA-B*4402	27*
K**E**ISNYTE**I**	HLA-B*4402	23*
R**E**INNYTN**L**	HLA-B*4402	25*
EMKLAITQT	HLA-A*26	15
Y**Y**IIYLEE**S**	HLA-A*2402	13
SQ**E**PSGKM**N**	HLA-A*01	10
K**L**AITQTY**Y**	HLA-B*1501 (B62)	22*
E**Q**DLLALD**K**	HLA-A*03	17
L**L**ALDKWA**S**	HLA-A*0201	15
L**L**ALDQWA**N**	HLA-A*0201	15
L**L**ALDSGQ**V**	HLA-A*0201	24*
L**L**ALASGQ**I**	HLA-A*0201	22*
Q**D**LLALDK**W**	HLA-B*4402	14
W**L**WIQDIH**N**	HLA-A*03	11
DTYRINSSS	HLA-A*26	20*
S**E**FGAALE**N**	HLA-B*4402	16
D**I**KIFIMI**V**	HLA-A*0201	21*

4.2.3 Epitopes for gp 41 Subtype C

A total of 39 epitopes identified from the conserved sequences of *gp 41* Subtype C were successfully analysed using SYFPEITHI. The analysis showed 12 epitopes with a score ≥20 and the best epitope having a score of 30 (Table 4.2). The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

Table 4.2: Conserved immunogenic epitopes for *gp 41* subtype C. The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

EPITOPE	HLA RESTRICTION	SCORE
G**A**LSSSRQ**E**	HLA-B*5101	14
G**H**LSSSRQ**E**	HLA-B*1510	13
CSGKLICT**T**	HLA-A*1101	11
G**I**WGCSGK**L**	HLA-A*0201	24
K**L**ICTTSV**P**	HLA-A*03	23
Q**L**TVWGIK**Q**	HLA-B*1501 (B62)	11
L**Q**TRVLAI**E**	HLA-B*1501 (B62)	11
R**Y**LKDQQL**L**	HLA-A*2402	25
G**I**WGCSGK**L**	HLA-A*0201	24
K**L**ICTTSV**P**	HLA-A*03	23
E**G**YRGATA**I**	HLA-B*5101	22*
ERYLKDQQL	HLA-B*1402	27*
P**R**GATAYS**S**	HLA-B*2705	13
A**Y**SSNSHS**R**	HLA-B*2705	13
QQHMLQLTV	HLA-B*1402	13
LTVYGIKQL	HLA-A*26	25*
H**I**VATHSL**G**	HLA-A*03	13
HSLGALSS**S**	HLA-A*1101	12
EGYRGATAY	HLA-A*26	21*
T**A**YSCNSR**S**	HLA-B*5101	17
L**G**LLWKTH**L**	HLA-B*5101	18
EIWETLTWM	HLA-A*26	26*
T**W**MQWERE**I**	HLA-B*5101	12
Q**D**IWDKMT**W**	HLA-B*4402	14
T**W**MQWDKE**L**	HLA-B*4402	13
I**E**KLIITQ**T**	HLA-B*4402	16
T**Q**TQYTSC**L**	HLA-B*2705	13
I**I**TQASYM**I**	HLA-A*0201	18
Q**A**SYMICL**K**	HLA-A*6801	18
EKDLLALDS	HLA-A*26	12

D **L** L A L D S W **K**	HLA-A*03	28*
R K **T** M **K** R I Y **Q**	**HLA-B*08**	**10**
R **I** Y Q H W Q L **E**	HLA-A*03	16
G Y **K** N **I** H N D **S**	**HLA-B*08**	**11**
K **I** C G V G L T **Y**	HLA-A*03	30*
Q T G C G Y K I F	**HLA-A*26**	**16**
I **Y** I F I M I V **G**	**HLA-A*2402**	**13**
F **I** M I V G G L **G**	HLA-A*03	14
K **P** V E L V C I **S**	**HLA-B*5101**	**14**

4.2.4 Epitopes for *gp* 41 Subtype D

A total of 22 *gp 41* Subtype A1 were identified and analysed using SYFPEITHI program. Out of the 22 epitopes identified, 3 epitopes with a score ≥20 and the best epitope had a score of 25 (Table 4.3). The epitopes with a score of greater than 20 were considered immunogenic and indicated by a red asteric.

Table 4.3: Conserved immunogenic epitopes for *gp* 41 subtype D. The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

EPITOPE	HLA RESTRICTION	SCORE
Y L **K** D **Q** Q L L **G**	**HLA-B*08**	**16**
I **W** G C S G K L **I**	HLA-B*5101	14
S G **K** L **I** C T T **A**	HLA-B*08	13
T **H** S L G H Y T **A**	HLA-B*3901	14
A **P** R N L G F S **G**	**HLA-B*0702**	**15**
I **N** S S S E F G **V**	**HLA-B*5101**	**11**
A **L** E N I F A P **L**	HLA-A*0201	22*
L T V W G I K Q L	HLA-A*26	25*
I **N** S S S E F G **V**	**HLA-A*0201**	**13**
A **L** E N I F A P **L**	HLA-A*0201	23*
I **N** G S S E F G **V**	HLA-A*0201	13
V **P** W N S S W S **N**	HLA-B*5101	13
E I W N N M T W M	HLA-A*26	15
T **W** M E W E R E **I**	**HLA-B*5101**	**14**
L **L** E E S Q T Q **Q**	HLA-A*03	15
L **I** E E S Q T Q **Q**	HLA-A*03	13
L **I** E A S K P S **R**	HLA-A*03	16
L **K** N R K P S K **K**	HLA-A*03	18

L**K**NRKSSK**K**	HLA-A*03	17
G**K**FVELVS**I**	HLA-A*0201	18
G**K**FVELVN**I**	HLA-A*0201	16
C**G**YKNIHN**D**	HLA-B*5101	12

4.3 Identification of Conserved Epitopes for *gp120* Sequences

In order to identify the conserved epitopes, amino acid sequences from the *gp* 120 region were aligned according to the different subtypes to which they belonged (Figure 4.10 a, b and c). The *env gp* 120 region sequences had some conserved sequences although some sequences could also not align due to variability in the amino acid sequences in this region (Figure 4.10). The different subtypes were aligned separately and conserved epitopes were selected from the conserved areas and are highlighted in a red colour. The identified epitopes were then analysed using SYFPEITHI to determine the immunogenicity of these epitopes. Epitopes from the different *gp 120* subtype alignment (Subtype A1, C and C) that were 9 amino acids long (9 mer long) were then analysed (Table 4.5-4.8) and given a score depending on their immunogenicity. Epitopes that had a score greater than 20 were considered immunogenic and are highlighted with red asteric. Figure 4.10 a, b and c show representative data from gp 120 subtypes A1, C and D alignment using ClustalW program and indicates the areas that are conserved.

a) *gp* 120 subtype D alignment

```
MADH012_         FLGVKPVVQLNYCWNGSLAEEDIVIRNCKSHKSCKKHNSTAS-CVCNNSLHKALQQYKTR
MYDH056_         ----KPVVSTQLLLNGSLAEEDIVIRNCKISQIIAKNIIVQLNESVTINCTRPYNNTRQG
STD788C15_       ---IKPVVSTQLLLNGSLAEEEIIIRSENITNNAKNIIVQLN-ESVTINCTRPSNNTRKG
STD788C11_       ---IKPVVSTQLLLNGSLAEEEIIIRSENITNNAKNIIVQLN-ESVTINCTRPPNNTRKG
STD788C1_        ---IKPVVSTQLLLNGSLAEEEIIIRTENITNNAKNIIVQLN-ESVTINCTRPSNNTRKG
STD788C6_        ---IKPVVSTQLLLNGSLAEEEIIIRSENITNNAKNIIVQLN-ESVTINCTRPSNNTRKG
STD788C7_        ---IKPVVSTQLLLNGSLAEEEIIIRSENITNNAKNIIVQLN-ESVTINCTRPSNNTRKG
STD788C2_        ---IKPVVSTQLLLNGSLAEEEIIIRTENITNNAKNIIVQLN-ESVTINCTRPFNNTRKG
STD788C12_       ---IKPVVSTQLLLNGSLAEEEIIIRTENITNNAKIIIVQLN-ESITITCTRPYQSQKRG
STD866C7_        ---IKPVVSTQLLLNGSLAEEEIIIRSENITNNAKNIIVQLA-TPVSINCTRPYNNTRKG
STD866C3_        ---IKPVVSTQLLLNGSLAEEDIIIRSENITNNAKTIIVQLT-EPVSITCIRPYNNTRKG
STD760C1_        ---IKPVVSTQLLLNGSLAEEEIIIRSENLTNNAKIIIVQLK-EAVTINCTRPYKSTRRG
TLHC009_         ---------TQLLFNGSFAEEDIIIRSENLTNNARTIIVHLH-KTLLLNCTRPNNNTRQG
TLHC021F_        ---------TQLLFNGSFAEEDIIIRSENLTNNARTILVHLH-KTFMLNCTRPNNNTRQG
MYDH057_         FLGVKPVVQLNYCWNGSLAEEDIVIRNCNLTNNAKNIIVQLN-ESVTINCTRPYNNTRQG
TLHC102_         ----------PLLLNGSLAEEDIIIRTANLTDNTQNIIVHLT-ATVMINCTRPNNNTRQG
TLHC040_         ----RSANSAINCCEWQSSRKRDNNSICKYHKQCQNYNSTSY-ESVTINCTRPNNNTRQS
MYDH062_         ---INQWCNSIARCMAVSQKKTYYLGSENLTNNAKTIIVHLN-KTVMINCTRPNNNTRQG
MADH007_         ---INQWCNSIARCMAVSQKKTYSLGSENLTNNAKTIIVHLN-KTVMINCTRPNNNTRQG
TKMH015_         ---------LHCCKWQSSRRRDNNSISKISQVILKTSYVQFN-ESITINCTRPNNNTRDG
TLHC023_         ----RSANSAINCCCMAVSQKKRSSLDLKISQTMPKLSSYILMSLSQLIVPGP-TTIQGK
```

b) *gp 120* Subtype C Alignment

```
TLHC001_         ---------NSTTAEWYVSLEQEIIITI--------------------------------
MYDH058_         -----SQWYQLNYCSWYVYAEEEIVIRISNLTNNAKTIIV--HLNESVEIVCTRTRSIIQ
STD796C16_       ---IKPVASTQLLLNGSLS-EEGIIIRSENLTNNAKTIIV--QFNKSVEITCTR-PNNNT
STD796C14_       ------------------------------------------------EITCPD-PPIIH
STD796C8_        ---IKPVVSTQLLLNGSLS-EEGIIIRSENLTNNAKTIIV--QFNKSVEITCTR-PNNNT
STD796C6_        ---IKPVVSTQLLLNGSLS-EEGIIIRSENLTNNAKTIIV--QFNKSVEITCTR-PNNNT
STD796C5_        ---IKPVVSTQLLLNGSLS-EEGIIIRSENLTNNAKTIIV--QFNKSVEITCTR-PNNNT
MYDH035_         ---SKPVVSTQLLLNGSLA-EEEIVIRISKSDRQCQNNNS--TASRPCKHCVVQGPNNNT
MYDH017_         ---DSGQWYQLNYCYGSLA-EDETIIRICICTN-NAIIIV--QLNKTVAINCTR-TQQYT
MADH015_         ---DNQWYQLNYSLNGTPSRKRDNNSISKICQTMPKQSIV--HLKTPVEIECIR-PGNNT
MYDH032_         ---DNQWYQLNYSLNGTPSRKRDNNSISKICQTMPKQSIV--HLKTPVEIECIR-PGNNT
MADH008_         -----SPVVSTQLLLMVVSQKKTSYLGSENLTNNAKTIIV--HLKDPVEIVCTRPNNNTR
MYDH009_         -----YPVVSTQLLLMVVSQKKTSYLGSENLTNNAKTIIV--HLKDPVEIVCTRPNNNTR
TLHC012F_        ---------STQLLLMVVSQKKRSSLGSENLTDNVKTIIV--HLQNPVEIVCTRPNNNTR
TLHC013_         ---------STQLLLMVVSQKKRSYLGSENLTDNVKTIIV--HLQNPVEIVCTRPNNNTR
MYDH031_         ------NRVSTHYCSMVVCRRKDS-NSSENLTNNVKLIIV--HLKKSVDIVCTRPNNNTR
MYDH036_         -----KPVVSTQLLSWSSRRRGDRNSGSKNLTDNAKIIIV--HLQDYVEIVCTRTRQYYK
TKMH017_         ----------SLLLNGSLAEEEQIIRS-ENLTDNTKTIIV--HLNESCRNSLYK-AQQYY
MYDH045X_        FFEMKPVVSTQLLLNGSLAEEDIVIRISKSGTDNAKIIIV--HLNESVRIVCTR-PGNNT
MADH002_         ----YASSINSTTVKWSSSRRRDSNSDLKICQTMPKYSSY--SLKT-LYTLCGTRAQQSY
MADH013_         -----KPVVSTQLLSWSSSRRGDSNSDLKICQTMPKSSSY--IFKT-MSRLCVQGPGNNT
TKMH018F_        -----SAICINSTTVNGSLAEEKIVIRISRSDRQCQNNNS--TSSRTLSKLCAQGPAIIL
TLHC012_         --------FQLNYCSMVRLAEERSYLGSKDLDRQCQNNNS--TSSRPCRNCVYTGPAIIL
MYDH019F_        ----INQWYQLSYYSNGSIAEDEIIIRIWKICKQCQNNNS--AASCTCRNKLYQDLTIIQ
MYDH013_         -------RINPLLLNGSLAEERGSNSSRRSWTNNRQQSMIGDILTESVRNCVYKAQQSNT
TLHC014_         --------STQLLLNGTFTQRETYSLDLKICQTMPKYYSY---SLMNLSQLIVCDPTIIQ
TKMH011F_        --------STQLLLNGTSSSNKRYYLRSENLTDNAKIMIV--QLKDPVEIECTETQQSYK
MYDH044_         ---IKPVVSTQLLLNGTSSRRRGGNYGLENLTNNAKTIIV--QLKKPSRNSLLQMQPSHF
MADH001_         ---RLSQCINSTTACMVVSQKKRSSLGSENLTNNAKTIIVTSYNPCSTVLLCITLRRGPR
MADH011_         --------------------CRDNNPDLKMCQTSAKTIIA--HLKEPVTNTSERDLAINT
MYDH2663_        ----AKPVYQLNYCMHGRLAEEEIVIRICKSAKQCQNNNSHILQPLSYCPSLHHTQEGTQ
MADH009_         -------RRAESTDSICEHYKQWPKTYSSNFTKHARNYVG----------TQTLTTNTST
MYDH018_         -------WGPSYLPAMFTICFILHGSVMVAFAVLLVVLIALLHLVYTCQQIRLYEQYAKF
TLHC018F_        ------LFGARLLLHMSIICFILHGSVDGFICITISTISPRMYHSTAMDLLRLLKMLLPQ
MYDH011_         ---------PCHVYNLFYSAWECRWYALVSLPVLLFSIWYY-YTRKDCYLYNKKPPATDK
MYDH043_         ---------FWGLPPLPFLSFVLFCIGVSWLHLKYYSYYSICMYCSTVLMYYNRKNSPPQ
MYDH010_         ----LLGARLPCTCLSFVLFCWECDGCISIIVCTIVSYYSTCMYYSTVWMCCNKKNSPLQ
TLHC024F_        ------CPFLSFVLFCIEVCWLNLKYYSYCYDMYVLLNSFGCIAICKILLHNSSYVLYSL
TLHC014F_        ---------LSLFAVFCIEELYVCINPSRSCETCVVLILWTPSAMLSDRFICYEDYSFTH
MYDH016_         -------PTQLSWYGSTVSRAAERIFPYQDWLMSWCFVLVDYSNCQGRVQGTFQSMCVTY
```

c) *gp 120* subtype A1 alignment

```
TLHC021F_       ------------------------------------ISQYYQLN--CCKWQSSRRRDKDS 22
TLHC004_        ------------------------------------ISQSYQLN--CCKWQSRRRRDKDS 22
MADH006_        -------------------------------------DQACSINSTAVEWHSSKRQGNNS 23
TLHC018F_       ------------------------------------------FNSTTVEWQFSRRRGKNS 18
TLHC007_        ------------------------------------------FNSTTVEWQFSRRRGKNS 18
TLHC037_        ------------------------------------FGTSQSFNTAVVKWQPSRKRGNNS 24
TKMH002_        -------------------------------------------------SMAVSQKERSC 11
MYDH036_        ------------------------------------------------------------
TLHC005_        ------------------------------------------VQLNCCKWQSSRRE-GKN 17
MYDH004_        ----------------------------------------AVVQLNCCYNGSLAKEKVKI 20
MYDH037_        ---------------------------------------IKPVVSTQLLYGSLAENQTKI 21
TLHC016F_       ---------------------------------------SQSFNSIAVKWQSSRK-GSNN 20
TLHC002_        ---------------------------------------SQSFNSIAVKWQSSRK-GSNN 20
MYSL032_        ---------------------------------------KPVVSTQLLLNGTLAKDRGNN 21
MYDH063_        -----------------------------------FFRLGQSVSTQLLLNGSLAKEEVKI 25
MADH010_        --------------------------------------TRQRASTQPLLNGSLAYEQLII 22
MYDHX5_A1       ------------------------------------------------------------
MYDH050_        --------------------------------------TSQCA-TQLLLNGSLAKGEVKI 21
TLHC007F_       ----------------------------------------ASSSTQLLLNGSLAEGKVII 20
TLHC017_        ----------------------------------------ASSSTQLLLNGSLAEGKVII 20
TLHC012F_       ----------------------------------------ASRSTQLLLNGSLAEGKVMI 20
TLHC033_        --------------------------------------------PIAVEWQSSRRRGNNS 16
MYDH046_        --------------------------------------KPVVSTQLLLNGTSSRRRDNNQ 22
MYDH039_        ----------------------------------FFGKRPVSINSAAAKWQSGQKIRQKL 26
L12_            ----------------------------------------------LCCWQLAEERNNS- 13
L25_            ------------------------------------------NSLLSMAVYQ---KERSC 15
TKMH019X_       --------------------------------------------LLSMAVYQ---QMGQK 13
MADH014_        ------------------------------------FLXTGPQYQLNYCKWQSRRRRGKN 24
MYDH008_        ---------------------------------------QASSINSTAVKWQSSKRRCNN 21
TLHC010F_       --------------------------------------CDQASSATNCCCMAVSQKKRSY 22
TLHC011F_       ----------------------------------------SQSFNSIRCQMAVQQKKRQK 20
TLHC024F_       ---------------------------------------TSQYFNSTAVEWQSSRRRSSN 21
TLHC036_        ---------------------------------------TSQYFNSTAVEWQSSRRRSSN 21
TKMH042F_       -------------------------------------FGQASSSTHCCSMESSRK-RSKN 22
TLHC003_        -------------------------------------FGQASSSTHCCYMESSRK-RSKN 22
TLHC011_        ----------------------------------------ASFSTQLLLNGSLAEGEIRI 20
MADH005_        ----------------------------------------AVVQLNCCYMAVSQKKRSRL 20
L26_            ---------------------------------------------LHCCKWQSSRRRGKD 15
TLHC0024_       ---------------------------------------PASSSTQLLVKWQPSRKRGKN 21
TKMH017F_       ----------------------------------------------TVVKWQPSRRRGNN 14
TLHC014F_       -----------------------------------FLGPASSSTQLLLNGSL-AEKEVII 24
MYDH027_        -----------------------------------FLGQARSINSTTVNGSL-AEGEVRI 24
MYDH033_        -------------------------------------DQACRINSTAVAMALSAKDKVII 23
L15_            ----------------------------------GFLGASAFCXXSTAAKWSSSXKGNSN 26
TLHC039X_       ---------------------------------------SAFCSTPLLLNGSLAEGKVKI 21
TLHC021F_       DLK-ISQTMPKTSSYNLTSLYA-LLVPDLTTIQGKVYCIGPG-QAFYATGDIIGDIRQAH 79
TLHC004_        ALK-ISQTMPKTYLYNLTSLYA-LLVPDLTTIQGKVYCIGPG-QAFYATGDIIGDIRQAH 79
MADH006_        DLK-ISQTMPKIYSYNLTSLCK-LLVSDLATIQEKVRRIGPG-QTFYATGDIIGDIR-AH 79
TLHC018F_       ICK--YHKQCQNHNSTIYCACKNSLYQDLTTIQGQVYCIGPG-QTFYATGDIIGDIRQAY 75
TLHC007_        ICK--YHKQCQNHNSTIYCACKNSLYQDLTTIQGQVYCIGPG-QTFYATGDIIGDIRQAY 75
TLHC037_        DLENISNNGAKIYLYNLPRLCT-LLVSDLITIQEKVYCIGPG-QTFYATDDVIGDIRQAY 82
TKMH002_        LGSANITNNAKNIIVQFTKPVKINCTRPNNNTR-KGIHIGPG-HAFYATDDIVGNIRDAY 69
MYDH036_        -----------------------------------GYDIGPRPNHFYAQEDIIGDIREAH 25
```

Figure 4.10: Conserved epitopes for *gp*120 subtypes alignment. Showing representative data for *gp 120* subtypes A1, C and D amino acid sequence alignment using ClustalW. Conserved epitopes highlighted in red colour

4.3.1 Conserved Epitopes for *gp* 120 Subtype D Sequences

A total of 51 epitopes were successfully analysed using SYFPEITHI and the analysis showed that *gp 120* Subtype D had 16 epitopes with a score ≥20 and the best epitope having a score of 28 as shown in table 4.4 below. Epitopes that were considered immunogenic had a score ≥20 are marked using red asterisks.

Table 4.4: Conserved Immunogenic Epitopes for Subtype D gp120.
The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

EPITOPES	HLA RESTRICTION	SCORE
N**G**SLAEED**I**	HLA-B*5101	17
R**P**SNNTRK**G**	HLA-B*0702	14
E**S**VTINCT**R**	HLA-A*6801	20*
EDIVIRNCK	HLA-A*26	16
L**L**NGSLAE**E**	HLA-A*0201	23*
I**K**PVVSTQ**L**	HLA-B*2705	16
I**K**PVVSTQ**L**	HLA-B*2705	17
L**L**NGSLAE**E**	HLA-A*0201	23*
EI**I**I**R**TEN**I**	HLA-B*08	20*
T**N**NAKNI *I **V**	HLA-B*5101	13
N**G**SLAEED**I**	HLA-B*5101	17
DIVIRNCNL	HLA-A*26	24*
E**S**VTINCT**R**	HLA-A*6801	20*
N**G**SLAEED**I**	HLA-B*5101	17
VTINCTRPN	HLA-A*26	12
R**P**NNNTRQ**S**	HLA-B*0702	12
R**P**YNNTRQ**G**	HLA-B*5101	15
R**C**MAVSQK**K**	HLA-B*2705	18
R**P**NNNTRQ**G**	HLA-B*0702	12
T**Y**YLGSEN**L**	HLA-A*2402	22*
T**N**NAKTII**V**	HLA-A*0201	14
K**T**VMINCT**R**	HLA-A*6801	19
R**P**NNNTRQ**G**	HLA-B*0702	12
N**G**SLAEED**I**	HLA-B*5101	17
DIVIRNCNL	HLA-A*26	24*
E**S**VTINCT**R**	HLA-A*6801	20*
R**C**MAVSQK**K**	HLA-B*2705	18
T**Y**YLGSEN**L**	HLA-A*2402	22*
T**N**NAKTII**V**	HLA-A*0201	14
R**P**NNNTRQ**S**	HLA-B*0702	12
R**P**NNNTRQ**G**	HLA-B*5101	11
K**T**VMINCT**R**	HLA-A*6801	19
T**R**ASTLYN**K**	HLA-B*2705	24*
K**T**STLYHS**W**	HLA-B*4402	14
T**H**IGPGRA**Y**	HLA-B*4402	17

FT**E**TITGD**I**	HLA-A*01	14
R**Q**AYCNIS**R**	**HLA-B*2705**	17
TSWNKTLQ**Q**	**HLA-A*1101**	14
VA**K**KLRDL**Y**	HLA-A*01	19
NTTKIIFKP	HLA-A*26	16
V**H**IGPGQA**L**	HLA-B*1510	23*
EITTHSFNC	HLA-A*26	15
S**G**LFNNSE**W**	**HLA-B*4402**	13
T**S**NSTSET**L**	HLA-B*3901	14
G**T**NESIIT**L**	HLA-B*3901	18
G**A**NEIIIT**L**	HLA-A*0201	24*
PC**R**I**K**QII**N**	**HLA-B*08**	16
F**I**AHGQMA**L**	HLA-A*0201	24*
C**R**IKQIIN**M**	HLA-B*2705	28*
F**I**AHGQMA**L**	**HLA-A*0201**	24*
N**L**TNNAKT**I**	HLA-A*03	18

4.3.2 Epitope for *gp* 120 Subtype C

Analysis of gp 120 using SYFPEITHI showed that subtype C had 16 epitopes with a score ≥20 and the highest score was 28 (table 4.5).

Table 4.5: Conserved immunogenic epitopes for Subtype C gp120. The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

EPITOPE	HLA RESTRICTION	SCORE
N**L**TNNAKT**I**	HLA-A*03	18
I**K**PVVSTQ**L**	HLA-B*2705	16
L**L**NGSLSE**E**	HLA-A*0201	21*
GI**I**I**R**SEN**L**	**HLA-B*08**	23*
T**N**NAKTII**V**	HLA-B*5101	12
KSVEITCT**R**	HLA-A*1101	18
DNQWYQLNY	HLA-A*26	19
S**L**NGTPSR**K**	**HLA-A*03**	27*
R**D**NNSISK**I**	HLA-B*2705	17
C**Q**TMPKQS**I**	HLA-B*1501 (B62)	11
KSVEITCT**R**	HLA-A*1101	18
DNQWYQLNY	**HLA-A*26**	19
YSLNGTPS**R**	HLA-A*1101	19
DNNSISKIC	HLA-A*26	11

Q**T**MPKQSI**V**	HLA-A*0201	18
T**P**VEIECI**R**	HLA-B*5101	13
S**T**QLLLMV**V**	HLA-A*0201	21*
SQ**K**K**T**SYL**G**	HLA-B*08	12
S**E**NLTNNA**K**	HLA-B*4402	15
T**N**NAKTII**V**	HLA-A*0201	14
H**L**KDPVEI**V**	HLA-A*0201	25*
H**L**QNPVEI**V**	HLA-A*0201	25*
CTRPNNNT**R**	HLA-A*1101	19
L**L**LNGSLA**E**	HLA-A*03	19
A**E**EEQIIR**S**	HLA-B*4402	17
A**E**EDIVIR**I**	HLA-B*4402	25*
WSSSRRRD**S**	HLA-A*1101	11
NS**D**LKICQ**T**	HLA-A*01	14
QTMPKYSSY	HLA-A*26	22*
D**R**QCQNNN**S**	HLA-B*2705	14
TSSRTLSKL	HLA-A*26	15
C**A**QGPAII**L**	HLA-B*5101	19
S**T**QLLLNG**T**	HLA-A*0201	17
S**I**CMYCST**V**	HLA-A*0201	23*
YS**Y**YSICM**Y**	HLA-A*01	20*
VS**Y**YSTCM**Y**	HLA-A*01	20*
S**T**CMYYST**V**	HLA-A*0201	19
R**T**SIRIGP**G**	HLA-A*6801	11
Q**T**FYATGE**I**	HLA-A*0201	16
IG**D**IREAH**C**	HLA-A*01	10
K**I**NGSEWN**E**	HLA-A*03	13
LK**E**HFPNK**T**	HLA-A*01	10
L**R**EHFPNT**T**	HLA-B*3901	13
K**S**IRIGPG**Q**	HLA-A*03	12
EKLKEHFPV	HLA-B*1402	14
EKLREHFPN	HLA-A*26	10
KSMRIGPG**Q**	HLA-A*1101	10
T**F**YATGEI**I**	HLA-A*2402	18
IS**E**E**K**WNE**T**	HLA-B*08	14
G**E**IRQAHC**N**	HLA-B*4402	16
R**K**LQEKFP**N**	HLA-B*2705	10
L**Q**EKFPNS**T**	HLA-B*1501 (B62)	10
G**E**LHKHFP**N**	HLA-B*4402	11
L**H**KHFPNK**T**	HLA-B*3901	12
T**R**DIIGDI**R**	HLA-B*2705	21*
SD**Q**D**K**HSM**Q**	HLA-B*08	12
T**G**DIIGNI**R**	HLA-B*2705	12
K**P**PSGGDL**E**	HLA-B*0702	13

V T T H S F N C **R**	HLA-A*1101	19
G **E** F F Y C N T **S**	HLA-B*4402	14
K **L** F N R T Y L **F**	HLA-B*1501 (B62	21*
G **T** Y N S N S N **I**	HLA-B*5101	16
S N I T L I T L P	HLA-A*26	10
A **P** S S G G A S **K**	HLA-A*03	23*
S **Q** H I A L I V **E**	HLA-B*1501 (B62)	11
E N F S I A I H Q	HLA-A*26	18
K **P** S S G G A S **K**	HLA-A*03	23*
S S G G A S K F **N**	HLA-A*1101	14
T **S** L S L S R R **I**	HLA-A*2402	13
F **L** L Q S S K L **F**	HLA-B*1501 (B62)	17
N **S** H T C L I V **I**	HLA-B*5101	17
H **A** N S T N R P **K**	HLA-B*5101	13
N S T N R P K F **T**	HLA-A*1101	11
C **R** I K Q I I N **M**	HLA-B*2705	28*
I **N** M W Q E V G **R**	HLA-A*6801	14

4.3.3 Epitopes for *gp* 120 Subtype A1

Subtype A1 had 42 epitopes with a score ≥20 and the epitope with the highest score had a score of 29 (table 4.6).

Table 4.6: Conserved Immunogenic epitopes for GP120 subtype A1. The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

EPITOPE	**HLA RESTRICTION**	**SCORE**
C **C** K W Q S S R **R**	HLA-B*2705	14
S S **R** R **R** D K D **S**	HLA-B*08	18
F S **R** R **R** G K N **S**	HLA-B*08	16
I **K** P V I S T Q **L**	HLA-B*2705	16
L **L** L N G S L A **E**	HLA-A*03	19
I **R** P V V S T Q **L**	HLA-B*2705	26*
L **L** N G S L A E **G**	HLA-A*0201	23*
S **L** A E G G K V **M**	HLA-A*03	22*
L **L** L N G S L A **K**	HLA-A*03	29*
S **Y** F P E I T H **I**	HLA-A*2402	23*
A **V** K W Q S S R **K**	HLA-A*03	27*
L **L** L N G T L A **K**	HLA-A*03	29*
T Q L L L N G S L	HLA-B*1402	18
S **L** A K G E V K **I**	HLA-A*0201	28*

A S S S T Q L L **L**	HLA-A*1101	20*
N **G** S L A E G K **V**	HLA-A*0201	12
S S **R** R **R** D N N **Q**	HLA-B*08	16
S S **R** R **R** G N N **S**	HLA-B*08	16
T S Q Y F N S T **A**	HLA-A*1101	11
V **E** W Q S S R R **R**	HLA-B*2705	16
F **G** Q A S S S T **H**	HLA-A*03	13
C **C** S M E S S R **K**	HLA-B*2705	15
F W **L** R A L L C **Y**	HLA-A*01	18
S S N T S S F L **D**	HLA-A*1101	16
N T S S F L D P S	HLA-A*26	13
I S Q T M P K T **S**	HLA-A*1101	12
S **Y** N L T S L Y **A**	HLA-A*2402	10
L **L** V P D L T T **I**	HLA-A*0201	28*
Q G **K** V **Y** C I G **P**	HLA-B*08	11
Q **A** F Y A T G D **I**	HLA-B*5101	21
Q **T** F Y A T G D **I**	HLA-A*0201	14
I **I** G D I R Q A **H**	HLA-A*03	17
I **I** G D I R Q A **Y**	HLA-B*1501 (B62)	19
I **R** S E N I T N **N**	HLA-B*3901	14
K **P** V K I N C T **R**	HLA-B*2705	18
C T R F N N N I **R**	HLA-A*1101	18
Q **A** F Y A T G D **I**	HLA-B*5101	21*
I T G D I R Q A S	HLA-A*26	12
I **R** S E N I T N **N**	HLA-B*2705	19
A **K** N I L V Q F **T**	HLA-A*0201	12
K **P** V K I N C T **R**	HLA-B*2705	18
C T R P N N N T **R**	HLA-A*1101	19
Q **A** F Y A T G D **I**	HLA-B*5101	21*
Q **T** F Y A T G D **I**	HLA-A*0201	14
I **I** G D I R Q A **H**	HLA-A*03	17
I **I** G K I R Q A **Y**	HLA-B*1501 (B62)	18
N **I** I G D I K Q **A**	HLA-A*0201	19
D I I G D I R Q A	HLA-A*26	21*
I **I** G D I K Q A **Y**	HLA-B*1501 (B62)19	19
I **I** G N I R E A **H**	HLA-A*03	17
I **R** S E N I S D **N**	HLA-B*3901	14
A K T I I V Q L **K**	HLA-A*1101	16
E P V N I T C T R	HLA-A*26	17
C T R P N N N T **R**	HLA-A*1101	19
A **Q** L T K A V N **I**	HLA-A*0201	15
M **I** C K Y H R Q **C**	HLA-A*03	11
R **I** G P G Q T F **Y**	HLA-A*03	24*
P **G** Q T F Y A T **R**	HLA-A*6801	14

C T **G** N I R Q A **Y**	HLA-A*01	22*
I **H** C F R P R N **N**	HLA-B*1510	15
P **R** N N T R T S **V**	HLA-B*3901	18
P **N** N N T R K S **I**	HLA-B*5101	12
D **L** E N I T N N **V**	HLA-A*0201	21*
K **N** I I V Q L V **N**	HLA-A*03	15
L **D** L K I S Q T **M**	HLA-B*2705	15
K **N** I I V Q L V **N**	HLA-A*03	15
R **L** Y K L I V P **D**	HLA-A*03	22*
V **P** D L T T I Q **E**	HLA-B*0702	12
T **M** L K P S L Y **T**	HLA-A*0201	16
S **D** L K I S Q I **M**	HLA-B*2705	14
P **K** L Y S Y N W **L**	HLA-B*2705	14
I **L** L I I N C T **R**	HLA-A*03	23*
C T R P N N N T **C**	HLA-A*1101	11
T S I L F V D Q **D**	HLA-A*1101	13
V D **Q** D **K** H S M **S**	HLA-B*08	11
D V I G D I R Q A	HLA-A*26	27*
V **I** G D I R Q A **H**	HLA-A*03	15
S **I** C K Y H K Q **C**	HLA-A*03	12
Y **H** K Q C Q N H **N**	HLA-B*1510	12
S **L** Y Q T S Q Q **Y**	HLA-A*03	24*
K **K** Y T V S D Q **D**	HLA-A*03	11
R **H** S M Q Q V T **Y**	HLA-A*03	18
S T T C P A C S **N**	HLA-A*1101	14
Y **L** Y Q T Y Q Q **Y**	HLA-A*03	23*
L **Y** Q D L T T I **Q**	HLA-A*2402	11
L **Y** Q A S Q Q Y **K**	HLA-B*2705	13
K **K** Y T V S D Q **G**	HLA-A*03	11
Y **H** S I Q Q V I **S**	HLA-B*3901	15
L **L** L R S Y S R **L**	HLA-A*0201	24*
P **L** L W S Y Q R **L**	HLA-A*0201	19
S **S** I S V L A S **L**	HLA-A*0201	22*
S **V** L A S L A S **A**	HLA-A*03	20*
S S N V A P V P **N**	HLA-A*1101	19
S **Q** G T F Q S M **C**	HLA-B*1501 (B62)	10
Q S M C V T Y L **H**	HLA-A*1101	12
L **P** Y Q E G A L **L**	HLA-B*5101	22*
D W T H P S C L L	HLA-A*26	17
T **H** P S C L L Q **C**	HLA-B*3901	17
C D **V** N **K** T E W **N**	HLA-B*08	10
I I F S S L L R **R**	HLA-A*1101	20*
N **I** Y S L L R E **G**	HLA-A*03	12
R **T** L Q Q V A K **Q**	HLA-A*6801	16

L**R**DKSLRI**K**	HLA-B*2705	24*
N**A**LGEVAT**Q**	HLA-A*6801	14
L**R**EHYFKN**K**	HLA-B*2705	22*
Q**E**VATQLR**E**	HLA-B*4402	11
T**I**IFNRSS**G**	HLA-A*03	17
GN**K**T**I**IFA**N**	HLA-B*08	13
I**I**FANSTG**G**	HLA-A*03	13
D**L**EITTHS**F**	HLA-B*1501 (B62)	18
I**T**THSFNC**M**	HLA-B*2705	11
D**L**EITTHS**F**	HLA-B*1501 (B62)	18
H**R**VATQLE**K**	HLA-B*2705	25*
HF**K**N**K**TII**F**	HLA-B*08	27*
I**I**FNSASG**G**	HLA-A*03	17
YF**K**N**K**TII**F**	HLA-B*08	26*
H**Q**VAEQLS**K**	HLA-A*03	17
N**Q**VAEQLS**K**	HLA-A*03	17
C**N**VSNSKW**K**	HLA-B*2705	12
ETLRQVAIQ	HLA-A*26	22*
R**Q**VAIQLR**K**	HLA-B*2705	21*
R**Q**VAIQLR**E**	HLA-B*1501 (B62)	13
D**L**EVTTHS**F**	HLA-B*1501 (B62)	18
VTTHSFNC**A**	HLA-A*1101	11
DIEITTHSF	HLA-A*26	21*
I**T**THSFNC**R**	HLA-A*6801	18
C**I**VNGSAW**N**	HLA-A*0201	12
K**A**LQQVAG**Q**	HLA-A*0201	14
Q**Q**VAGQLG**K**	HLA-B*2705	16
IF**V**N**K**TIR**F**	HLA-B*08	17
I**R**FDKPSG**G**	HLA-B*2705	16
C**N**LNKTTW**H**	HLA-B*2705	13
A**T**LPKGSQ**P**	HLA-A*03	18
K**L**YQILRR**G**	HLA-A*03	17
S**R**NNNTSF**S**	HLA-B*3901	11
WI**K**M**E**SNF**T**	HLA-B*08	14
KM**E**SNFTT**G**	HLA-A*01	10
WI**K**M**E**YNF**T**	HLA-B*08	14
I**V**TSVRQN**G**	HLA-A*03	14
I**E**LYNRSL**H**	HLA-B*4402	14
S**L**HNSENT**S**	HLA-A*03	16
S**K**SQHIVL**I**	HLA-A*0201	16
S**Q**HIVLIV**E**	HLA-B*1501 (B62)	12
Q**S**IRMESH**F**	HLA-A*03	15
Q**S**IRMESD**F**	HLA-A*0315	15
Y**L**LTPQEW**I**	HLA-A*0201	23*

Y **L** L T P Q E G **V**	HLA-A*0201	26*
Y **L** L T P Q E G **I**	HLA-A*0201	24*
R **R** I F L L C S **S**	HLA-B*2705	19
S S L F N S T W **M**	HLA-A*1101	14
G **E** F F Y C N T **S**	HLA-B*4402	14
T **S** S L F N S I **W**	HLA-B*4402	12
S G **D** S T E S G **D**	HLA-A*01	11
L **I** I L P C R I **K**	HLA-A*03	22*
Q **I** I N M W Q R **V**	HLA-A*0201	20*
E **N** I T L P C R **I**	HLA-B*4402	15
K **Q** I I N M W Q **R**	HLA-B*2705	16
I **R** M W Q R T G **Q**	HLA-B*2705	13
S T Q G A N S T **D**	HLA-A*1101	14
I **I** T L P C R I **K**	HLA-A*03	21*
Q **I** I N M W Q R **V**	HLA-A*0201	20*
G **E** F F Y C N H **Q**	HLA-B*4402	13
G **E** F F Y C N H **Q**	HLA-B*4402	13
Q A C L I A L G L	HLA-B*1402	19
L **C** R R G F S I **V**	HLA-A*0201	14
E **W** H C S M Q E **L**	HLA-B*1510	13
I **H** G R Y T L H **A**	HLA-B*3901	14
E S F S I V I H Q	HLA-A*26	22*
I **H** Q A C L I A **L**	HLA-B*3901	25*

4.4 Identification of Conserved Epitopes for the *Gag* Region

Sequences from the *gag* region were also aligned using CLUSTALW program to determine their level of conservation. The different subtypes were aligned separately and a representative data of the aligned sequences indicated by figure 4.11a, b and c. The analysis showed that these sequences from gag region showed a high degree of conservation among the different clades (figure 4.11 a, b and c). Conserved epitopes were selected and highlighted in a red colour.

4.4.1 Conserved Epitopes for *Gag* Sequences

Conserved epitopes from *gag* region were analysed using the SYFPEITHI program. Epitopes from subtype A1, D and C were analysed. From gag region, Q R L W S D R V L epitope was highly conserved in gag subtype A1 and is an HLA-

B*1402 epitope with a score of 25. The epitopes with the highest scores were selected to be used in vaccine construction. The SYFPEITHI program predicted that the epitope K V L G K R L C K (from gag subtype D; table 4.7) is highly promiscuous (restricted to more than one HLA) and it is recognised highly in the context of HLA HLA-A*03 with a score of 32. This was followed by K I C G V G L T Y an HLA-A*03 epitope from gp 41 subtype C (table 4.9).

a) *Gag* Subtype D alignment

```
MADH007_             ---------GFSPCQHFGHKTRTKGNPLETMSIG-SIKLSEP-SKLHRRSKTGCQKPCWS 49
MYLD005_             ---------GFSPCQHFGHKTRTKGNPLETMYIG-SIKLSEP-SKLHRRSKTGCQKPCWS 49
TKMH019_             --------QHFG------HKTGTKG-TLSRLCRS-VLSNSKSRASFTGCKKLDDRNLVGS 44

TKMH011_             ------------APSILDIRQGPKE-PFRDYVDR-FYKTLRAEQATQGGKKLDDCNLVGP 46
MYSL032_             ------------ALQHFGHKTGAKG-PFRDYVGPGFYKTLRAEQASQDVKNWDDRNLVGS 47
MYDH008X_            -----GLVR-FSPCQHFGHKTRTKG-TLSRLRRS-VLSNSETEQASQDVKNWMTETLL-V 51
MADH001_             -----GLVR-FSPCQHFGHKTRTKG-TLSRLRRS-VLSNSETEQASQDVKNWMTETLL-V 51
MADH012_             -----------LALPAFWTYDKAKG-TLSRLCRS-VLYNSKSRASFTSMYKIGCQKPCWS 47
L8_                  -----------LALPAFWTSDKAKG-TLSRLCRS-VLSNSKSRASFTSMSKIGCQKPCWS 47
MYDH025_             -----SGLVRFTALQHFGHKTRTKG-PFRDYVGS-VLSNSKSCASFTRCKNLDCQKPCWV 53
MYDH057_             -----------LGPPAFWTSNKAKG-PFRRLCRIGSIKLSEP-SKLHRMSKIGCLETLLI 47
MYDH022_             PTCHFQNSCDVSPCSTLDIRQGPRN-PLETMVGP-VLSNSKSRASFTRWKTWMTETLL-G 57
TLHC014_             ---LLGXSKFXGRPSILGHKTRTKXTLXEDYVRS-ALSNSKSCAMXTXMQXIRMTDTF-- 54
```

```
MADH007_             KMLNPDCKTILKALGPALHSKKCCQHVRE-WGDPVIKQGFWLRQCAKQQISILLS---CC 105
MYLD005_             KMLNPDCKTILKALGPALHYKKCCQHVRE-WGDPVIKQGFWLRQCAKQQISILLS---CC 105
TKMH019_             KMQIQIVKLSSKHWDQGLHSKKCCQHVKE-WEDPVIKQGFWLRQCAKQQMQMLLY---CC 100
TKMH011_             KCEPRLYEYLKSIWGPAATLEEMMTACQG-VGGPSHKARVLACGNEPSIKCKCCY---ND 102
MYSL032_             RMLNPDCKTILKALGPVQPHSKKCCQHVR-EWGAQPSSKSFGCGNEPSIRCKCCY---ND 103
MYDH008X_            QNANPDCKTILKALGPGATLGEMMTACQG-VGGPAYG-KGFGCGN-CAKQQVGLL---SC 105
MADH001_             QNANPDCKTILKALGPGATLGEMMTACQG-VGGPAYG-KGFGCGN-CAKQQVGLL---SC 105
MADH012_             KMLNPDCKTILKALGPAATLEEMMTACQG-VGGPSHKARVLAEAMSQATQANTAV---MM 103
L8_                  KMLNPDCKTILKALGPAATLEEMMTACQG-VGGPSHKARVLAEAMSQATQANTAV---MM 103
MYDH025_             QNANPDCKTILKALGPGATLGRHDDSMSGESGGPSHKSKEFLACGNCSQATNGKCXSCCC 113
MYDH057_             QNAEPRLSNYLKSTVFRGYIGGKCCQHVKEWGGPSHKARVFGCGNCAKSIPPA------- 100
MYDH022_             PKCEPRLSNYLKSIGTRGYIGGTCCQLVSXXXGPSYSSKXXRXEXXXXXXIGXAX----- 112
TLHC014_             --XGPKCXPRLXXLSS-------------------------------------------- 68
```

b) *Gag* subtype A1 alignment

```
MADH011_             ----SSPCQHFGHKA-----RAKCT-LQKIMSDR-FFKTL--RAKQATQEVKGCIT-ETL 46
MYLDL03_             ----SSPCQHFGHKA-----RAKCT-LQKIMSDR-FFKTL--RAKQATQEVKGCIT-ETL 46
TKMH012_             ---FSQTVRFIALLAFWISNKGPKN-HSEIMYDR-FFKAL--RAEQATQEVKHCMT-ETL 52
MYDH037_             ---------LYSPVSILDIKQGPKE-PFRDYVDR-FFKVL--RARASHTGGKRLDDMKHC 47
MADH005_             --SFSDLVRFIRPVSILDIKQGPKE-PFRDYVGSGSLKLS--ELSKLHRRSKVGWT-ETL 54
MYDH028_             ----------DSPVSILDIRQGPKE-PFRDYVRSVLYNSQ--SLSKPHRMSKIGST-ETL 46
MYLD07_              -------------VSILGHKTRAKR-PFRDYVDTVLSNSQ--SCASYTGGKRLGCQ-KPC 43
MYDH038_             -------------VSTLDIKQGPKE-PFRDYVDRILSNSQ--SCASYTRGKRMDRL-ETL 43
MADH010_             ----------PSGAQHFGHKTRAIC-NPSETMSTGSLKLS--EPNKLHKRSKLGST-ETL 46
L9_                  ----------PSGAQHFGHKTRAIC-NPSETMSTGSLKLS--EPNKLHKRSKLGST-ETL 46
TLHC023_             ---------GSLKLSEL---SKLHR-RYK--VG--------------MTG--------NI 23
MYSL033_             -------TSQILALQHIGYKTRAKR-PFRDYVGQ-VLSSS--QSCASHTGCKKLDD-RNI 48
TKMH010x_            --------------SHFGCKTRAKR-PFREYVDT-VF----------------------- 21
MYDH063              --------SLYMPSSHFGYKARSKR-PFRDYVDR-FFKTL--RARASYT-GGKRLD--CQ 45
```

```
MYDH0035_     ---------FLRPASILGYKTGAKR-PFRGLCRIGFFKTL--RSLKQATQGGKKXG--XT 46
MADH014_      ---------RCIALLASWISNKGQK-NPSEIMSIGSLKLS--ELSKLHRRSK-VGCQKHY 47
MYDH110_      ---------RCIALLASWISNKGQK-NPSEIMSIGSLKLS--ELSKLHRRSK-VGCQKHY 47
MADH004_      ---------RFLALLAFWISNKGQK-YPSEIMSIGSLKLS--ELSKLHRKSK-VGCQKHY 47
L2_           ---------RFLALLAFWISNKGQK-YPSEIMSIGSLKLS--ELSKLHRKSK-VGCQKHY 47
MYDH026-1_    ---LFNNSSDVSPCSHFGYKTRAKR-PFRDYVGSGSLKFS--ELSKLHRMSK-IGCQKHY 53
MYDH026-2_    ---LFNNSSDVSPCSHFGYKTRAKR-PFRDYVGSGSLKFS--ELSKLHRMSK-IGCQKHY 53
MYDH049_      -----------AP-SHFGYKTRAKR-TLQSIMSIGSLKFS--EPSKLHRRSK-VGCQKHY 44
MYLDX005_     -----------FSPASILVSNKGQK-NLSETMSIGSLKLS--ELSKLHRMSK-IGCQKHY 45
TLHC019_      ----FIRPASILDIK------QGPK-NPSEIMYDRVLQSP--QSCTSFTGGKRLDCQTHY 47
MYDH006_      ----FVRTSKMLALLVFWISNKGQK-NPSEIMSDRVLSNS--QSCASYTGGKRLGCQKHY 53
TLHC018_      -----------IALLAFWISNKGQK-NPSEIIVDRFFKVL--RAEQATQVGKRFGCQTHY 46
MYDH046_      -------SYVVYALLAFWISNKGQK-NPSEIMSIWFFKVL--RAEQATQEVKGSDRQKHY 50
MADH008_      ------PSKIFSPASILGYKNRGQK-TLQRLWSDRVLLKSQ-SCASYTGGKGVGXDKXLL 52
MYDH015_      ----LLKPVRFSPCSFLD-SNRGQK-TLQRLWWDRVLSNSQ-SLSKLHREVKGSDDRKHY 53
MADH003_      ----FLITVRLSPCSHFGYKTRAKR-TLQRLCR--------------------------- 28
TLHC033_      -----------SPCYHFGHKTRAKR-TLQRLCRS-VLSSSQ-SCTSYTGCEKLDDMKHYW 46
MYSL031_      ---------IFYPCYHFGYKTMAKR-PFICLCIYGSLKLSC-SGSSYTWMVKWVGWTDTL 49
MADH002_      --PPFSCLVRFIALLAFWISNKGQK-TPSEIMVGSDSLKLS-ELEQATQEVK-GWMTETL 55
MYDH019F_     --PPFSCLVRFIALLAFWISNKGQK-TPSEIMVGSDSLKLS-ELEQATQEVK-GWMTETL 55
MYSL030_      ----FLRTSKILALLVFWISNKGHC-TLQRLCR-SGSLKLS-ELEQATQEVK-GWMTETL 52
MYDH022X_     PLFLKTIVSFYSPPSILGYKHRAHKDPSGLYVAQGSSKALK-TLTYAPTGMSKNLGIPEK 59
MYDH055_      ----FYDLVRFYALLAFWISNKGQR-TLSEIMSDRVLSGFSGACASSTGMSKKLGCQKPL 55
MYLD08_       ----------------------FQK-NPSETMYDMVLSNS--QSCACYTGRYKVRMTETL 35
```

c) *Gag* subtype C alignment

```
MADH009_      ---------GVSPAAFWKYDKGQRNPLET-MSDRFFKTLR-AEQATQE-VKNWMTETLLV 48
MYDH111_      ---------GVSPAAFWKYDKGQRNPLET-MSDRFFKTLR-AEQATQE-VKNWMTETLLV 48
MYDH004_      ----------SALAAFWTSDKVQKNPLET-MYDRFFKTLR-AEQATQD-VKNWMTETLLV 47
MYDH016_      ---FVNNSSDVSPAAFWTSNKGQKNPLET-MSDRFFKTLR-AEQATQD-VKNWMTDTLLV 54
TKMH017_      --------------QHFGHKTRAKGPFRD-YVDRFFKTLR-AEQATQE-VKNWMTDTLLV 43
MYDH007_      ------------ALQYFGHKTRAKDHLRD-YVDRFFKTLR-AEQATQD-VKNWMTDTLLV 45
MYDH0020_     -GFLMEQSGCLALSAFWDIKQGPKEPFRRLWYDRFFKTLRSWNKLHKM-SKIGWTDTLLV 58
MYDH024_      ------------------------DYVR-----PVLSNLKSCTSYTSM-YKIWMTDTLLV 30
MADH015_      ---------RLYSPAAFWTSDKDQRNLLETMSTGSLK-LYELNKLHKM-SKIGCQTPCWS 49
TLHC120_      ---------RLYSPAAFWTSDKDQRNLLETMSTGSLK-LSELNKLHKM-SKIGCQTPCWS 49
TLHC012_      ----------FYRPSSILDIKQGPKDPLETMSTGSLK-FYELNKLHKK-SKIGCQTPCWS 48
TKMH010_      ---------SIYSPAAFWTSNKGQRTFYGLCRPVLSN-IIELNKLHKR-SKIGCQTPCSV 49
MYDH040_      ----LFKASEVYSPAAFWTSNKGQRSPLETMSTGSLK-PSELNKLHKM-SKIGCQKPCYV 54
MYDH012_      ------------GPAAFWHKTRSKGTFSRLCRSVLSN-FKQLNKLHKM-SKIGCQTPCSV 46
MYDH039_      -------LAFCVAPQHFGHKTRTKRTLSRLCTTDSLQ-LSELNKLHKM-SKIGCQTPCSV 51
MYDH018_      -----------LALQHFGHKTRAKGPFRDYVCPVLSN-LKSCTSYTRC-QKLDGQTPCWS 47
MYDH0021_     -------NSKMLALQHFGHKTRAKEPFRDYVGPVLSN-LKSCTSFSRC-KKLDGQTPCWV 51
MYSL1030_     -----------RPRQHFGHKTRAKEPFRDYVDRVLSN-LKSLNKLLKM-YKIGCQTPCWV 47
L10_          --------------------KG-QMTFSRLCRPVLSN-LKSCTGYTRC-KKLDDRHLVG- 36
MYDH013_      ------PSKFVYPRSILDIKQGPKEPFRDYVRPVLYN-FKSGTSYTRC-KKLDDRHLVG- 51
MYDH062_      --------RGLWTLQHFGHKTRAIGTLSKTSGXGSLKPSRSCTSYTXMXKNWDDXKPWXV 52
MYDH002_      ----FYSLVRFIALQHFGHKTRTKGTLSRLCRTGSFKP---------------------- 34
MYDH023_      ------------APQHFGHKSGSKVTLXXLSYDRFFKTLSSCTSYTRX-KKLDDXTPCAV 47
MYDH036_      ----------FSPSSILDIKQGPKDPLEIMWT-GFSN-LKELEQAFQDVKKLGCETPLLG 48
MYDH027_      VKQGVLKTGRFYPFIIFDMKQWPNGPFIYYGRPGLSN-FKDLEQSTQDGKNLMNMTPFS- 58
MYDH047_      ------------GPAAFWTSNXXQXNPLXTMSTGSLKLSE-QNKLHKI-VKNWMTXHLVG 46
MYDH052_      -----FGAYYFGGPVAFWTSNKVQKNLLETMSDRFFKTLR-AEQATQD-VKNWMTDTLFG 53
MYDH009_      ------------SPCSILDIKQGPKDPLETMYDRFFKTLR-AEQATQD-VKNWMTGYLVG 46
MYDH0056_     -----LLGQCDFSPAAFWTSDKAKGPFRD-YVDRFFKTLRAEQATQEV--KNWDDRHLVG 52
MYDH001_      ----FLINRRMYSPAAFWT----------------------------------------- 15
MYDH011_      ----------------FWGEKPAXXP---------HLKLSSGQATQDV--KNWIDRHLV- 32
MYDH048_      ----PPFEHSRFSPPAFCQRQG-LGPFRD-----LGPVLKTQSCTASQDVKNWIQTLLGX 50
MYDH003_      ---------WILPLQHFGNKTRPNYPLET-MSDRFFKTLRSWNKLHRE-VKIWMTETLVG 49
```

Figure 4.11: Conserved epitopes for *Gag* Sequences. Showing representative data for *gag* subtypes A1, C and D amino acid sequence alignment using ClustalW. Conserved epitopes highlighted in red colour

4.4.2 Conserved Epitopes for *Gag* Subtype D

Analysis using SYFPEITHI showed that subtype D had a total of 14 epitopes with a score ≥20 and the highest (epitope K **V** L G K R L C **K)** had a score of 32, followed by R **R** P L A R N C **K** with 26, G P **L** S **R** K C R **L** with 25, C A **K** Q **Q** V G L **L** with 25, Q **A** T T G N A A **I** with 23, Q **R** G N F K G Q **K** with 22 **21, and** W **E** D P V I K Q **E** with 26. (table 4.7). A total of 66 sequences were analysed using the SYFPEITHI program.

Table 4.7: Conserved immunogenic epitopes for *Gag* subtype D sequences. The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

EPITOPE	**HLA RESTRICTION**	**SCORE**
G **F** S P C Q H F **G**	HLA-B*2705	6
N **P** L E T M S I **G**	HLA-B*5101	16
L N **K** I **V** R M Y **S**	HLA-B*08	10
P V S I L D I K **Q**	HLA-A*1101	14
S **I** L D I R Q G **P**	HLA-A*0201	14
F **G** H K T R T K **G**	HLA-B*5101	12
F S P C Q H F G **H**	HLA-A*1101	10
Q **H** F G H K T G **A**	HLA-B*1510	12
F **G** H K T G A K **G**	HLA-B*5101	11
L A **L** P A F W T **Y**	HLA-A*01	19
P **A** F W T S N K **A**	HLA-B*5101	15
S K **L** H **R** R S K **T**	HLA-B*08	10
T G **C** Q **K** P C W **S**	HLA-B*08	10
V **L** S N S K S R **A**	HLA-A*0201	15
A **S** Q G V K N W **M**	HLA-B*2705	14
F Y **K** T **L** R A E **Q**	HLA-B*08	12
D V K N W D D R N	HLA-A*26	18
D V K N W M T E T	HLA-A*26	18
K **M** L N P D C K **T**	HLA-A*0201	18
H S **K** K **C** C Q H **V**	HLA-B*08	10
Q **N** A N P D C K **H**	HLA-B*2705	12
I **L** K A L G P Q **A**	HLA-A*03	21*
T **L** E E M M T A **C**	HLA-A*0201	12
Q N A N P D C K T	HLA-A*6802	10

I**L**KALGPA**A**	HLA-A*03	19
T**L**EEMMTA**C**	HLA-A*03	13
E**E**MMTACQ**G**	HLA-B*4402	15
K**M**LNPDCK**T**	HLA-A*0201	18
EP**R**L**S**NYL**K**	HLA-B*08	16
KTILKALGP	HLA-A*26	16
WG**D**PVIKQ**G**	HLA-A*01	11
W**E**DPVIKQ**E**	HLA-B*4402	20*
K**P**QMEMLL**Y**	HLA-B*4402	13
V**G**GPSHKA**R**	HLA-B*2705	12
V**L**AEAMSQ**A**	HLA-A*0201	19
Q**A**TGSNGA**I**	HLA-B*5101	23*
Q**A**TTGNAA**I**	HLA-B*5101	22*
S**Q**QQGQMV**P**	HLA-B*1501 (B62)	12
Q**Q**QGQMVP**N**	HLA-B*1501 (B62)	10
C**G**NEPSIK**C**	HLA-B*5101	11
PS**I**KCKCC**Y**	HLA-A*01	19
CA**K**Q**Q**VGL**L**	HLA-B*08	25*
V**G**GPSHKA**R**	HLA-B*2705	12
V**L**AEAMSQ**A**	HLA-A*0201	19
Q**A**TQANTA**V**	HLA-B*5101	20*
R**E**AILEAK**R**	HLA-B*2705	18
Q**R**GNFRGP**K**	HLA-B*2705	20
ESLSVSTV**A**	HLA-A*1101	13
R**R**PLARNC**K**	HLA-B*2705	26*
K**V**LGKRLC**K**	HLA-A*03	32*
RSIKCFNC**G**	HLA-A*1101	12
K**E**GHLARN**C**	HLA-B*4402	13
L**A**RNCRAP**R**	HLA-B*5101	12
R**I**IKCFNC**G**	HLA-B*45	15
Q**R**GNFKGQ**K**	HLA-B*2705	22*
Q**R**GNFKGP**K**	HLA-B*2705	21*
QSGNFKAQ**G**	HLA-A*1101	11
C**R**KAFSGA**K**	HLA-B*2705	21*
KSIKCFNC**G**	HLA-A*1101	12
K**E**GHTAKN**C**	HLA-B*4402	14
T**A**KNCRAP**R**	HLA-A*6801	20*
EKLSMFTGQ	HLA-A*26	11
GP**L**S**R**KCR**L**	HLA-B*08	25*
R**H**LARYCR**V**	HLA-B*3901	17

4.4.3 Epitopes for *Gag* Subtype A1

A total of 80 epitopes from the conserved areas on gag subtype A1 were analysed using SYFPEITHI. *Gag* subtype A1 had a total number of 16 epitope with a score ≥20 with the highest score of 25 (Table 4.8).

Table 4.8: Conserved Epitopes for Gag subtype A1 sequences. The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

EPITOPE	**HLA RESTRICTION**	**SCORE**
L N **K** I **V** R M Y **S**	HLA-B*08	10
P V S I L D I K **Q**	HLA-A*1101	14
S S P C Q H F G **H**	HLA-A*1101	13
S **P** V S I L D I **K**	HLA-A*6801	14
V S I L G H K T **R**	HLA-A*1101	19
L **G** H K T R A K **R**	HLA-B*2705	15
R **C** I A L L A S **W**	HLA-B*4402	14
S **W** I S N K G Q **K**	HLA-A*03	18
R **F** L A L L A F **W**	HLA-B*4402	12
F **W** I S N K G Q **K**	HLA-A*03	17
S **H** F G Y K T R **A**	HLA-B*1510	14
F **G** Y K T R A K **R**	HLA-B*2705	17
A **L** L V F W I S **N**	HLA-A*03	17
F **W** I S N K G Q **K**	HLA-A*03	17
H F **G** Y **K** T R A **K**	HLA-B*08	13
F **G** Y K T R A K **R**	HLA-B*2705	17
P **F** R D Y V R S **V**	HLA-B*5101	13
V **R** S V L Y N S **Q**	HLA-B*3901	13
M S T G S L K L **S**	HLA-A*1101	13
N **P** S E T M S T **G**	HLA-B*5101	16
N **P** S E I M S I **G**	HLA-B*5101	13
Y **P** S E I M S I **G**	HLA-B*5101	14
V **G** S G S L K F **S**	HLA-B*5101	11
T **L** Q R L W S D **R**	HLA-A*03	20*
Q **R** L W S D R V **L**	HLA-B*1402	25*
R **A** E Q A T Q E **V**	HLA-B*5101	18
T Q **E** V **K** N W M **T**	HLA-B*08	12
E L S K L H R R S	HLA-A*26	11
E L S K L H R K S	HLA-A*26	12
E L S K L H R M S	HLA-A*26	12

E**L**EQATQE**V**	HLA-A*0201	20*
L**V**QNANPD**C**	HLA-A*03	11
L**I**QNANPD**C**	HLA-A*0201	11
L**G**PKCKSR**L**	HLA-B*5101	18
T**G**PECKSR**L**	HLA-B*5101	17
LSVNFKSI**R**	HLA-A*1101	19
C**S**KMQIQI**V**	HLA-A*0201	13
S**P**FSEHSD**Q**	HLA-B*5101	11
S**P**FSEHSE**Q**	HLA-B*5101	11
W**S**KMQIQI**V**	HLA-A*0201	13
N**P**FYEHSD**Q**	HLA-B*5101	13
I**G**ATLEEM**M**	HLA-B*2705	10
T**A**CQGVGG**P**	HLA-B*5101	13
GH**K**A**R**VLA**E**	HLA-B*08	21*
T**A**CQGVGG**P**	HLA-B*5101	13
SH**K**A**R**ILA**E**	HLA-B*08	22*
GL**H**S**K**RCC**Q**	HLA-B*08	18
GH**K**A**R**VLA**K**	HLA-B*08	21*
GH**K**A**R**VLA**E**	HLA-B*08	21*
HA**R**E**W**EDP**A**	HLA-B*08	13
AIKQGCWL**R**	HLA-A*1101	19
A**I**RQGFWL**R**	HLA-A*03	20*
AIKQGFWL**R**	HLA-A*1101	20*
A**M**SQVQHP**N**	HLA-A*0201	11
I**M**MQRGNF**K**	HLA-A*03	17
I**M**MQRGNF**R**	HLA-B*2705	14
IMMQRATL**R**	HLA-A*1101	17
V**Q**NTNIMM**Q**	HLA-B*1501 (B62)	12
M**Q**RGNFRG**G**	HLA-B*1501 (B62)	11
I**M**MQKGNF**R**	HLA-B*2705	15
A**I**SQVQGA**S**	HLA-A*03	14
A**M**SQVQQT**N**	HLA-A*0201	12
QK**R**I**K**CFN**C**	HLA-B*08	16
QK**K**I**K**CFN**C**	HLA-B*08	20*
S**E**KDSVFQ**L**	HLA-B*4402	24
W**Q**KRTPSQ**K**	HLA-A*03	21
L**Q**APVGKR**L**	HLA-B*1510	15
VEKELSVST	HLA-B*18	15
IK**G**L**K**RFN**C**	HLA-B*08	12
R**E**VILEAK**E**	HLA-B*4402	11
G**L**SVSTVA**K**	HLA-A*03	25*
T**V**AKKGTS**P**	HLA-A*03	18
R**K**AILGGQ**K**	HLA-A*03	23*
K**I**KCFNCG**K**	HLA-A*03	20*

E**G**HLARNC**R**	HLA-A*6801	13
V**L**GKRAVK**C**	HLA-A*03	19
K**I**AGPYEK**G**	HLA-A*0201	16
GP**Y**E**K**GCE**N**	HLA-B*08	20*
Q**C**ARSAYK**H**	HLA-B*2705	13
G**R**HWEGMM**D**	HLA-B*2705	15

4.4.4 Conserved Epitopes for *Gag* Subtype C

Subtype C had a total of 11epitopes with a score ≥20 and the highest had a score of 32 as shown in table 4.9 below.

Table 4.9: Conserved Immunogenic Epitopes for Gag subtype C. The epitopes with a score of greater than 20 were considered more immunogenic and indicated by a red asteric.

EPITOPE	HLA RESTRICTION	SCORE
P**A**AFWKYD**K**	HLA-A*6801	15
K**G**QRNPLE**T**	HLA-A*03	10
PVSILDIK**Q**	HLA-A*1101	14
Q**G**PKEPFR**D**	HLA-B*5101	10
R**L**YSPAAF**W**	HLA-A*03	24*
T**S**DKDQRN**L**	HLA-B*1510	14
SRLCRSVLS	HLA-B*1402	20*
L**E**TMSTGS**L**	HLA-B*4402	21*
Q**H**FGHKTR**A**	HLA-B*1510	14
K**G**PFRDYV**C**	HLA-B*5101	10
DYVCPVLSN	HLA-A*26	18
L**K**SCTSYT**R**	HLA-B*2705	13
L**E**TMSDRF**F**	HLA-B*4402	21*
SDRFFKTL**R**	HLA-A*1101	14
M**S**DRFFKT**L**	HLA-B*4402	16
YVDRFFKTL	HLA-A*26	19
V**D**RFFKTL**R**	HLA-B*2705	13
V**K**NWMTET**L**	HLA-A*0201	16
I**G**CQTPCW**S**	HLA-B*5101	10
I**G**CQTPCS**V**	HLA-B*5101	18
K**K**LDDRHL**V**	HLA-A*0201	15
K**L**DDRHLV**G**	HLA-A*03	23*
L**E**EMMTAC**Q**	HLA-B*4402	11
L**E**EMMTSC**Q**	HLA-B*4402	11
T**I**RRNDDS**M**	HLA-A*03	16
S**G**SGRPGH**K**	HLA-A*03	17
RP**G**H**K**ARV**L**	HLA-B*08	26*
HF**K**S**I**RTR**G**	HLA-B*08	11

G **V** G R P G H K **A**	HLA-A*03	16
G **V** G G P G H K **A**	HLA-A*0201	15
R **V** L A E A M S **Q**	HLA-A*03	24*
A **E** A M S Q M N **N**	HLA-B*4402	13
A **E** A M S Q L N **N**	HLA-B*4402	14
L **A** E A M S Q A **N**	HLA-B*5101	14
P **R** R I V K C F **N**	HLA-B*2705	12
C **G** K E G T S P **K**	HLA-A*03	17
G **T** S P K I A G **S**	HLA-A*6801	16
P **R** R N I K C F **N**	HLA-B*2705	12
L **Q** G P R K K G **G**	HLA-B*1501 (B62)	11
G P **R** K **K** G G G **N**	HLA-B*08	23*
A **E** A L S Q V N **N**	HLA-B*4402	15
A **E** A M S Q V N **N**	HLA-B*4402	14
G H **K** A **R** V L A **E**	HLA-B*08	21*
S H **K** A **R** V L A **E**	HLA-B*08	22*
C **Q** H V R E W E **D**	HLA-B*1501 (B62)	10
H F **K** S **I** R T R **G**	HLA-B*08	11
V **A** R K A P S S **K**	HLA-A*03	19
K **V** L G K R L C **K**	HLA-A*03	32

The regions of an individual epitope that were most immunogenic are bolded and the silent regions that do not elicite a strong immune response are not bolded. The regions underlined were not very immunogenic. Some epitopes in this study were less immunogenic hence gave a lower score when presented with particular HLA type but more immunogenic when presented with a different HLA type for example I **M** M Q R G N F **K** an A1 epitope from *gag* region gives a score of 17 when presented with HLA-A*03 while when presented with HLA-B*2705 gives a score of 14. If the most immunogenic region of an epitope (bold) is altered or replaced with a different epitope, the immunogenicity of the entire epitope is interfered or reduced for example I **M** M Q R G N F **K** epitope (table 4.8).

4.5 Location of the Conserved Epitopes in the HIV Genome

In order to successfully analyze the generated epitopes in this study, the exact location of these epitopes was determined. A total of 167 conserved immunogenic epitopes that had a score greater or equal to 20 were all selected and their location in the HIV-1 full genome determined. This was done so that these epitopes can be compared with other epitopes that have already been identified previously.

Determination of their exact location and comparison with the others already identified also helps in the establishment if the epitopes have not been identified before. The locations for the different epitopes from the three different regions that is Gag, env gp120 and gp 41 were done separately (table 4.10-12). Using the BLAST tool, location of the immunogenic epitopes was established. Tables 4.11- 4.13 indicate the locations of the epitopes in the HIV genome. These are epitopes from *gp* 120, *gp* 41 and *gag* regions generated in this study.

Table 4.10: Location in the HIV genome of the 41conserved immunogenic epitopes identified from *gag* in this study.

Epitope	Location in the HIV Genome
T**L**QRLWSD**R**	1244-1260
Q**R**LWSDRV**L**	1230-1254
E**L**EQATQE**V**	1253-1274
GH**K**A**R**VLA**E**	1879-1905
SH**K**A**R**ILA**E**	1106-1131
GH**K**A**R**VLA**K**	1861-1883
GH**K**A**R**VLA**E**	1879-1905
A**I**RQGFWL**R**	1064-1089
AIKQGFWL**R**	1064-1089
QK**K**I**K**CFN**C**	2195-2219
S**E**KDSVFQ**L**	2195-2219
W**Q**KRTPSQ**K**	2706-2772
G**L**SVSTVA**K**	2700-2724
R**K**AILGGQ**K**	1349-1374
K**I**KCFNCG**K**	1310-1334
GP**Y**E**K**GCE**N**	1250-12 74
I**L**KALGPQ**A**	1247-1273
W**E**DPVIKQ**E**	1061-1086
Q**A**TGSNGA**I**	1099-1125
Q**A**TTGNAA**I**	1099-1125
CA**K**Q**Q**VGL**L**	1247-1273
Q**A**TQANTA**V**	1310-1334
R**R**PLARNC**K**	1369-1385
K**V**LGKRLC**K**	1345-1361
Q**R**GNFKGQ**K**	1126-1151
Q**R**GNFKGP**K**	1246-1271
C**R**KAFSGA**K**	1205-1230
T**A**KNCRAP**R**	1634-1660
GP**L**S**R**KCR**L**	1680-1704
R**L**YSPAAF**W**	1630-1654
SRLCRSVLS	1685-1709
L**E**TMSTGS**L**	1705-1730
L**E**TMSDRF**F**	1700-1724
K**L**DDRHLV**G**	1156-1181
RP**G**H**K**ARV**L**	1053-1078
R**V**LAEAMS**Q**	1416-1442
GP**R**K**K**GGG**N**	1081-1103
GH**K**A**R**VLA**E**	1204-1230
SH**K**A**R**VLA**E**	1106-1131
SH**K**A**R**VLA**E**	1286-1311
R**I**KCFNCG**K**	1150-1176

Table 4.11: Location in the HIV genome of 86 conserved immunogenic epitopes from clades A1, D and C identified from *gp 120* region.

Epitope	Location in the HIV Genome
E**S**VTINCT**R**	6572-6597
L**L**NGSLAE**E**	6198-6224
EI**I**I**R**TEN**I**	6407-6433
DIVIRNCNL	6575-6597
E**S**VTINCT**R**	6572-6597
T**Y**YLGSEN**L**	6500-6524
T**R**ASTLYN**K**	6540-6574
V**H**IGPGQA**L**	6737-6761
G**A**NEIIIT**L**	6205-6227
F**I**AHGQMA**L**	6594-6620
C**R**IKQIIN**M**	6844-6870
F**I**AHGQMA**L**	6594-6620
L**L**NGSLSE**E**	6500-6524
GI**I**I**R**SEN**L**	6454-6479
S**L**NGTPSR**K**	6400-6424
S**T**QLLLMV**V**	6176-6198
H**L**KDPVEI**V**	6353-6379
H**L**QNPVEI**V**	6500-6524
A**E**EDIVIR**I**	6211-6234
QTMPKYSSY	6200-6224
S**I**CMYCST**V**	6200-6230
YS**Y**YSICM**Y**	6302-6319
VS**Y**YSTCM**Y**	6340-6372
T**R**DIIGDI**R**	6300-6324
K**L**FNRTYL**F**	6845-6864
A**P**SSGGAS**K**	6737-6761
K**P**SSGGAS**K**	6700-6724
C**R**IKQIIN**M**	6844-6870
I**R**PVVSTQ**L**	6685-6706
L**L**NGSLAE**G**	6996-7019
S**L**AEGGKV**M**	6210-6236
L**L**LNGSLA**K**	5533-5560
A**V**KWQSSR**K**	5523-5547
L**L**LNGTLA**K**	6549-6575
S**L**AKGEVK**I**	5520-5540
ASSSTQLL**L**	5533-5547
L**L**VPDLTT**I**	6700-6750
Q**A**FYATGD**I**	6760-6786
DIIGDIRQA	6720-6754
R**I**GPGQTF**Y**	6356-6382
CT**G**NIRQA**Y**	7186-7207

D**L**ENITNN**V**	7100-7124
R**L**YKLIVP**D**	5934-5955
I**L**LIINCT**R**	6718-6732
DVIGDIRQA	6366-6399
S**L**YQTSQQ**Y**	6400-6432
Y**L**YQTYQQ**Y**	6314-6399
L**L**LRSYSR**L**	6400-6450
S**S**ISVLAS**L**	6500-6524
S**V**LASLAS**A**	6500-6530
L**P**YQEGAL**L**	6600-6630
IIFSSLLR**R**	6700-6750
L**R**DKSLRI**K**	7295-7315
L**R**EHYFKN**K**	7289-7315
H**R**VATQLE**K**	6500-6528
HF**K**N**K**TII**F**	6688-6714
YF**K**N**K**TII**F**	7215-7239
ETLRQVAIQ	6380-6405
R**Q**VAIQLR**K**	6429-6455
DIEITTHSF	6621-6646
Y**L**LTPQEW**I**	7234-7250
Y**L**LTPQEG**V**	6431-6449
Y**L**LTPQEG**I**	6431-6449
L**I**ILPCRI**K**	6920-6943
Q**I**INMWQR**V**	6670-6696
I**I**TLPCRI**K**	7218-7254
Q**I**INMWQR**V**	6670-6696
ESFSIVIHQ	6806-6832
I**H**QACLIA**L**	6538-6563
I**R**PVASTQ**L**	6685-6708
L**L**NGSLAE**E**	6254-6280
L**L**NGSLAE**G**	6216-6242
EIIGDIRQA	6645-6671
EVVKQLGKH	6486-6502
N**L**TLPCRI**K**	7122-7150
C**R**IKQIIN**M**	6844-6870
DRVARKLSV	6850-6870
SSHLANST**K**	6839-6864
A**L**SQWNRL**Y**	6749-6767
Y**P**MQISNV**I**	6780-6802

Table 4.12: Location in the HIV genome of 37 cross-clade conserved immunogenic epitopes identified from *gp 41* region.

Epitope	**Location in the HIV Genome**
G**I**KQLQAR**V**	7357-7382
G**I**WGCSGK**L**	7320-7354
LTIYAQHHL	6900-6924
A**M**MPMLWE**M**	6832-6854
I**P**KWSTAP**M**	7320-7324
EFRGTLGKL	7594-7620
G**I**WGCSGR**L**	7800-7824
GVINLKRRY	7392-7418
K**R**NLAITQ**I**	6589-6605
K**E**IDNYTQ**L**	7576-7600
K**E**ISNYTE**I**	7269-7295
R**E**INNYTN**L**	7287-7309
K**L**AITQTY**Y**	7354-7370
L**L**ALDSGQ**V**	7410-7443
L**L**ALASGQ**I**	7354-7378
DTYRINSSS	6629-6655
D**I**KIFIMI**V**	7705-7730
G**I**WGCSGK**L**	7205-7228
K**L**ICTTSV**P**	7204-7228
R**Y**LKDQQL**L**	7152-7178
G**I**WGCSGK**L**	7400-7424
K**L**ICTTSV**P**	7300-7324
E**G**YRGATA**I**	7323-7345
ERYLKDQQL	7390-7415
LTVYGIKQL	6621-6694
EGYRGATAY	7299-7318
EIWETLTWM	7488-7514
D**L**LALDSW**K**	7534-7559
K**I**CGVGLT**Y**	7500-7524
A**L**ENIFAP**L**	7270-7293
LTVWGIKQL	6890-6908
A**L**ENIFAP**L**	7270-7296
DTSRINSSS	6629-6655
ETHGLGHKQ	6600-6624
K**E**IGNYSD**I**	4976-5000
K**E**IGNYTD**I**	7326-7352
EIWENMTWL	7271-7297

4.6 Design of Cross-Clade Multiepitope Candidate Vaccine

In this study, epitopes with highly significant scores (above 25) using SYFPEITHI were selected and used to develop a candidate DNA vaccine that is 300 amino acid long. These are epitopes that have been identified are conserved in subtype A1, D and C subtypes of HIV. The epitopes were put together in order to construct a candidate vaccine is cross-reactive to protect against the major clades circulating in Kenya. Epitopes that were selected for vaccine design are from *gag*, *gp120* and *gp41* regions, with scores of 25 and above. The designed region specific super-epitope are shown in figure 4.12.a below. The gaps in between indicate the links between the different epitopes.

Gag:
Q **R** L W S D R V **L-**G **L** S V S T V A **K-**C A **K** Q **Q** V G L **L-**R **R** P L A R N C **K-**
K **V** L G K R L C **K-**G P **L** S **R** K C R **L-**R P **G** H **K** A R V **L-**K **V** L G K R L C **K**

Gp120:
C **R** I K Q I I N **M-**S **L** N G T P S R **K-**C **R** I K Q I I N **M-**E V V K Q L G K H-
Y **L** L T P Q E G **V-**Y F **K** N **K** T I I **F-**H F **K** N **K** T I I **F-**D V I G D I R Q A-
L **L** V P D L T T **I-**S **L** A K G E V K **I-**L **L** L N G T L A **K-**L **L** L N G S L A **K-**
C **R** I K Q I I N **M-**

Gp41:
G V I N L K R R Y-K **R** N L A I T Q **I-**K **E** I D N Y T Q **L-**R **E** I N N Y T N **L-**
R **Y** L K D Q Q L **L-**E R Y L K D Q Q L-L T V Y G I K Q L-E I E T L T W M W-
D **L** L A L D S W **K-**K **I** C G V G L T **Y-**L T V W G I K Q L-E I W E N M T W L-

Figure 4.12 a: Region specific combined epitopes. Showing conserved epitopes from subtype A1, C and D from the three regions of HIV-1 namely *gag, gp120* and *gp 41.*

All these region specific combined epitopes were then linked together to give a super-epitope of about 300 amino acids long. The resulting super-epitope is shown in figure 4.12.b below:

Q **R** L W S D R V **L** G **L** S V S T V A **K** C A **K** Q **Q** V G L **L** R **R** P L A R N C **K**
K **V** L G K R L C **K** G P **L** S **R** K C R **L** R P **G** H **K** A R V **L** K **V** L G K R L C **K**
C **R** I K Q I I N **M** S **L** N G T P S R **K** C **R** I K Q I I N **M** E V V K Q L G K H
Y **L** L T P Q E G **V** Y F **K** N **K** T I I **F** H F **K** N **K** T I I **F** D V I G D I R Q A
L **L** V P D L T T **I** S **L** A K G E V K **I** L **L** L N G T L A **K** L L N G S L A **K**
C **R** I K Q I I N **M** G V I N L K R R Y K **R** N L A I T Q **I** K **E** I D N Y T Q **L**
R **E** I N N Y T N **L** R **Y** L K D Q Q L **L** E R Y L K D Q Q L L T V Y G I K Q L
E I E T L T W M W D **L** L A L D S W **K** K **I** C G V G L T **Y** L T V W G I K Q L
E I W E N M T W L

Figure 4.12.b: 300 amino acid super epitope construct.
The epitopes from *gag, gp41* and *gp120* that had a score ≥25 from subtypes A1, C and D were pooled together.

4.7 Epitopes Identified Previously By Other Investigators

Cytotoxic T lymphocytes (CTL) restricted epitopes from *gag p24* region, *gp120* and *gp 41* regions identified by other investigators (tables 4.13, 4.14 and 4.15) were selected and used to compare with the epitopes that was generated in this study. The epitopes identified in this study were from a similar location in the HIV genome with the epitopes that were have been previously identified by other investigators. A total of 43, 13 and 12 epitopes from gag, gp41 and gp120 respectively from other investigators were used for comparison in this study.

Table 4.13: *Gag* region-p24 epitopes identified by other investigators and their location in the HIV genome. (Sridharan et al., 2005; Janssens et al.; 1997)

Epitope	Location in the HIV Genome
QAISPITLNAW	1462-1487
ETINEEAAEW	942-971
RPNNNTRKSI	6577-6606
GEIYKRWII	992-1017
DCKTILKAL	1199-1225
DLNTMLNTV	761-787
DRFYKTLRA	1227-1253
KRWIILGLNK	840-869
GHQAAMQML	1321-1343
TSTLQEQIGW	1510-1539
PMFSALSEGATPQDLNTM	1159-1190
GATPQDLNTMLNTVGGH	1329-1379
WMTNNPPIPVGEIYKRWI	739-792
PVGEIYKRWIILGLNKIV	1102-1155
GPKEPFRDYVDRFYKTLR	1135-1187
YVDRFYKTLRAEQASQEV	1153-1205
LRAEQASQEVKNWMTETL	1127-1176
TPQDLNTML	1320-1346
VIPMFSAL	839-861
ACQGVGGPGHK	1265-1297
RDYVDRFFKTL	1689-1721
EVIPMFSAL	714-740
SPRTLNAWV	1464-1490
EIYKRWII	765-787
GLNKIVRMY	1129-1155
RRWIQLGGLQK	1863-1888
KRWIILGLNK	1576-1701
NPVPVGNIY	1818-1844
PPIPVGDIY	1092-1118
SEGATPQDL	1326-1349
EAQASQDVKNW	1151-1181
RMYSPTSI	1606-1628
QASQEVKNW	1139-1165
ISPRTLNAW	1462-1487
KAFSPEVIPMF	1472-1503
KAFSPEVI	1275-1298
TSTLQEQIGW	1510-1539
TSTVEEQIQW	952-981
TPQDLNTML	13230-1326
VIPMFSAL	1352-1366
TPQDLNTML	1320-1346

DLNTMLNTV	1528-1553

Table 4.14 Envelope *gp* 41 epitopes identified by other investigators and their location in the HIV genome. (Manfred et al., 1997; Hilleman et al.; 2001)

Epitope	Location in the HIV Genome
SLLNATDIAV	6575-6595
RLRDLLLIVTR	5494-5525
YLKDQQLL	7384-7407
RGPGRAFVTI	7161-7178
IPRRCRQGL	6666-6684
IVTRIVELL	7766-7788
ERYLKDQQL	7178-7208
RYLKDQQLL	7152-7178
GRRGWEALKY	7988-8011
HIPRRIRQGLERALL	8366-8410
TAVPWNASW	7512-7538
QELKNSAVSL	8044-8070
RAIEAQQHL	7898-7924

Table 4.15: Envelope *gp* 120 epitopes identified by other investigators and their location in the HIV genome

Epitope	Location in the HIV Genome
TVYYGVPVWK	5898-5927
LFCASDAKAY	5731-5769
RIKQIINMW	6686-6912
RPNNNTRKSI	6577-6606
VPVWKIATTTL	6263-6285
VPVWKEATTT	6335-6363
SFNCGGEFF	6775-6805
TVYYGVPVWKEATTTLF	6037-6086
YRLISCNTSVITQACPKV	6245-6297
RVKEKYQHL	387-403
DPNPQEVL	6457-6477

CHAPTER FIVE: DISCUSSION

5.1 The Study Overview

In this study, a total of 139 sequences of the HIV-1 subtypes circulating in northern Kenya were generated mainly from the *env gp41*, *env gp120 (C2V3)* and the *P24 gag* regions of the HIV genome. These are the main regions that are targeted in design of HIV DNA vaccines. Eventually, it is hoped that a multi-epitope vaccine can be generated that incorporates all the different subtypes of HIV-1 circulating in northern Kenya. Subtype specific epitopes were identified that is for HIV-1 subtype A, C and D which are the common subtypes of HIV-1 circulating in Kenya. In line with this, different subtypes of HIV-1 were used to generate epitopes and then aligned together so as to come with a 'consensus' epitope (s) by use of bioinformatics tools in attempting to design a multi epitope cross-clade HIV-1 vaccine to reflect viral diversity. These 'cross-clade epitopes' may be the critical elements required for developing an effective HIV-1 vaccine enhancing existing vaccines given the extent of HIV-1 variability and the global context of HIV.

One of the greatest barriers to developing an effective HIV -1 vaccine is the diversity of viral strains and clades. The genetic variation of microorganisms and their evolution in time has important implications for control of infectious diseases. Genetic variation may be reflected in the differences in biological characteristics that may determine transmissibility, pathogenesis and immunogenicity (Zolla-Pazner *et al.,* 1998). Genetic variability of microorganisms needs to be taken into account when developing or adapting diagnostic tests and vaccines and when making projections of the burden of morbidity and mortality.

Identification of strains or subtypes has also proved to be an invaluable tool in studying the spread of infectious pathogens ((McMichael *et al.,* 1990). There is insufficient data regarding the HIV-1 subtype distribution on the Kenyan borders. Understanding these dynamics would be useful in diagnosis, treatment, and prevention and in the continued search and evaluation of candidate vaccines. In Africa HIV-1 diversity is the highest in the world. All the subtypes have been found although A1 and C subtypes are the most prevalent (Janssens *et al*., 1999), the greatest being in west and Central Africa (McCutchan *et al*., 1999; McCutchan, 2000). The number of unique HIV –1 sequences in public databases has been steadily increasing (Altschul *et al*., 1997; De Groot *et al.,* 2001)

5.2 HIV-1 subtype diversity in Northern Kenya

The three border districts (Turkana, Moyale and Mandera districts) were used to determine the subtypes of HIV circulating in Northern Kenya. The region borders several countries: Ethiopia, Somalia and Sudan. Partial sequencing of *gp 41* sequences and phylogenetic analysis showed that 44% of the sequences generated from the three districts were subtype A1, 45% were subtype C and 10% were subtype D. From Moyale 36% of the samples were subtype A1, 55% were subtype C and 9% were subtype D. Subtype C is the predominant subtype in the Moyale region because of the cross border movement from Ethiopia.In Ethiopia, the dominant HIV subtype in circulation is HIV-1 subtype C, based on partial genome sequencing (Abebe *et al.,* 2001).

From Mandera 67% were subtype A1, 33% were subtype C. There was no subtype D in this region. In Turkana the most dominant subtype was A1 (58%), subtype C was 25% and subtype D was 17%. Turkana borders Uganda and Sudan but little is known about the HIV epidemic in Sudan. However, few studies show that Sudan seems to be dominated by East African subtypes A, C and D (Matthew *et al.*, 2002). Phylogenetic analysis from 30 samples collected in 1998 and 1999 in Sudan showed that the dominant subtype of HIV-1 in circulation was D (50 %) C (30 %) and the rest HIV-1 subtype B and A (McCarthy *et al.*, 1995). In Uganda, the predominant subtypes are A and D (Hu *et al.*, 2000). This means that the distribution patterns of HIV-1 subtypes that are currently circulating in the world are probably the result of virus exchanges between populations (viral recombination), rather than mutations or diversification within different human populations (Myers, 1994).

The results from this study showed that in Northern Kenya (Mandera and Turkana districts) subtype A1 is the dominant subtype except in Moyale where subtype C is the most predominant. Previous studies carried out in Northern Kenya by Khamadi *et al.* (2005) also indicated that subtype A1 was the most dominant except in Moyale, which was dominated, by subtype C (48%). Presence of recombinant subtypes such as A1C, A1CD and A1D, and that the predominant subtypes circulating in Northern Kenya may have been influenced by cross border movements (Khamadi *et al.*, 2005).

The results also indicate that there is a higher proportion of HIV-1 subtype C circulating in northern Kenya, as compared to the rest of the country. Though data on HIV-1 subtype distribution in northern Kenya is scanty due to limited research carried out there. In other parts of Kenya, circulating subtypes include A1, C and D with A1 being the most predominant (Dowling *et al* 2002). A full-length subtype G has also been reported previously from Kenya (Carr *et al.,* 1998), with recombinants between A1, A2, and D; A2 and D; A1 and D; A1 and G; A1 and C; A1, C, and D; and C and D (Dowling *et al.,* 2002). This suggests that recombination plays an important role in the genetic diversification of HIV-1. On this basis the diversity in other parts of Kenya is probably due to a consequence of superinfection and / or intersubtype recombination.

5.3 Identification of Conserved Epitopes

Los Alamos national laboratory has shown that the number of unique HIV –1 sequences in public databases has been steadily increasing, with no end in sight. This is because of the high diversity of HIV due to evolution of the virus, due to mass movements of people from different areas and also due to recombination (Gaschen *et al*., 2002; Khamadi *et al.,* 2005). Although HIV-1 diversity is a problem, studies have shown that many conserved regions are still present in the different clades and therefore this gives hope in construction of vaccines.

In this study, generated amino acid sequences from the three regions (*gp 41, gp 120*, and *gag)* were successfully aligned using CLUSTAW program version 1.81 (Thomson *et al*., 1994) to identify the conserved amino acid epitopes. The sequences from the three regions (*gp* 120, *gp* 41 and *gag*) were aligned differently according to the different subtypes present, that is, subtype A1, C and D. None of the vaccines currently under evaluation contains the whole of the HIV genome. Therefore use of consensus or conserved sequences for the design a vaccine could provide a partial solution to the problem of HIV-1 sequence diversity since conserved epitopes increase the breadth of induced immune response. Studies have been carried out to establish this hence it is viable and can be utilized in cross-clade vaccine design (Gaschen *et al*., 2002).

The amino acid sequence alignment for *gp* 41 sequences showed that there was conservation of amino acid sequences among the different clades but some could not align due to variability in the amino acids. The *env gp 120* region was also conserved in almost all subtypes but some sequences could also not align due to variability in the amino acid sequences. This is because of the high amino acids variability in the envelope region. The conserved epitopes were selected from the conserved areas of the amino acid sequences. For vaccine development this amino acid variability in the envelope region may be disadvantageous.

The changes in the amino acid sequence that are associated with HIV-1 diversity may interfere with ligand processing and binding in the MHC binding groove and may diminish cross-clade protection against HIV –1 challenge by T cell clones (Paradela *et al*., 2000). This is because sequence modifications may affect the intracellular processing of the epitopes prior to MHC binding, by affecting processing by proteasome or transport of the protein through TAP into the endoplasmic reticulum (Altivia and Margalit, 2000; Paradela *et al*., 2000). Alternatively, variant peptides that still bind to the MHC may fail to engage the TCR, acting as 'antagonistic' to T cell response (Klenerman, 1994). Changes in the sequence of HIV-1 peptides that conforms to the binding specificities of a particular MHC can compromise binding of the modified peptide to that MHC (Zevering and Khamboonruany, 1994).

Modifications of MHC ligands at the amino acid level have also been associated with failure to bind or to be recognized by the T cell resulting in viral escape from the immune response (McMichael and Philips, 1997; Goulder *et al*., 1997; Evans *et al*., 1999). Conserved proteins from different subtypes share high amino acid homologies and hence more closely related than variable protein from the same subtype and this might be exploited by using a single vaccine strain for conserved proteins and multiple clade-specific strains for variable vaccines specifically designed to target these conserved epitopes (Gaschen *et al*., 2002).

This study also established that the *gag* region was extensively conserved among the different clades since most of the sequences were able to align as compared to *gp* 120 and *gp* 41 sequences. This could suggest that this region of the HIV genome may be ideal for design of a HIV vaccine due to the high level of conservation noted despite the different clades. In order to overcome or contend with variability in HIV it would be advisable to use the gag region. The vaccine developed from the conserved epitopes from this study would protect not only against the clades in Northern Kenya but also in other parts of Kenya. Therefore identification of conserved, immunologically relevant cross-clade epitopes may be the critical element required for developing an effective HIV-1 vaccine and enhancing existing vaccines (Gaschen *et al*., 2002).

5.4 Identification of Immunogenic Epitopes Using SYFPEITHI

Most vaccines today are created to generate either killer T-cell (CTL) or antibody responses. Although most vaccines have not been able to elicit a strong antibody neutralizing reaction that is cross reactive among all the HIV-1 clades, focus has been shifted to the induction of CTL responses (McMichael and Rowland-Jones, 2001). The ability of a vaccine to induce responses directed at a particular group of epitopes is of interest because it should be easier to access the possible efficacy of any resultant immune responses. Therefore the knowledge of epitopes is therefore critical in the precise evaluation of the strength and quality of CTL responses that could be induced by vaccine candidates.

In this study, the epitopes that were conserved were selected for further evaluation using epitope prediction offered by SYFPEITHI database of MHC ligands and peptides (Rammensee *et al.,* 1999), epitopes with the highest probability of association with HLA alleles were selected. A total of 41 conserved immunogenic epitopes (9 amino acids in length) from gag region were identified, 80 epitopes from gp 120 and 37 epitopes from gp 41) and their HLA restriction indicated.

Selection of peptides that were highly likely to bind to MHC enabled the identification of immunogenic cytotoxic T-cells (CTLs) epitopes, which are critical components of HIV vaccines. SYFPEITHI indicates silent regions of epitopes that elicit poor or no immune response. The epitopes that were not immunogenic were given a score of up to negative one (-ve 1) and the best epitope obtained had a score of 32. In this study, epitopes that had a score of less than 10 were considered poorly immunogenic while the epitopes with a score of greater or equal to 20 were considered immunogenic.

From this study, it would be helpful to identify immunologically silent regions of a vaccine so that these can be omitted from future constructs. For example in this study, altering the most immunogenic (Bolded) areas alters the immunogenicity of an epitope, for example gp 41 A1 subtype epitope, Q **A** F Y A T G D **I** with a score of 21 was altered (Q A **F** Y **A** T G D **I)** and the score reduced to 14. Another epitope G **V** G R P G H K **A** had a score of 16 but when some of the immunologically silent regions (underlined) of the epitope was altered (G **V G G P G H K A)** immunogenicity was not so much altered (score of 15).

Most of the identified conserved immunogenic epitopes from this study were HLA-A*26, HLA-B*4402, HLA-A*0201, HLA-A*03, HLA-B*1402, HLA-B*5101, HLA-A*2402, HLA-B*0702, HLA-A*1101, HLA-B*2705, HLA-B*3901, HLA-B*08, HLA-A*01, HLA-A*2402. The most common HLA was HLA-A*03, HLA-A*26, HLA-B*08 and HLA-A*0201. .

Characterization of HIV-1 specific responses has shown that populations differ in their HLA allele frequencies hence the HIV-1 specific responses detected differ significantly. Mapping of minimal epitopes would allow the design of gene-based vaccines based on locally prevalent HLA types that could cover specific regions for practical reasons

(Goulder *et al*., 2001. Therefore potential vaccines need to be tailored not only based on local viral sequence diversity, but also on the local HLA allele distribution.

The current understanding of HLA supertype is incomplete (Brander, 2004). The HLA molecule may have a more pronounced influence on disease progression than the epitope (Brander, 2004). For example HLA –B57 and B27 have been associated with slower HIV disease progression (Carrington *et al*., 1999). However the role of the presented epitope and the restricting HLA allele in the disease progression is still unclear (Brander, 2004). Recent data also indicates that HLA-B156 and HLA –B 1517 which fall into the same B 58 HLA supertype as HLA-B 57 may also be associated with slow HIV disease progression (Brander, 2004). Individuals expressing HLA-B 1516 and B 57 may provide beneficial effect. Other alleles such as A3 are also associated with disease progression (Frahm *et al*., 2002). The identification of these HLA supertypes facilitates the epitope-based vaccine development in that the selection of promiscuous peptide binders to those alleles included within a given supertype limits the number of peptides to be considered without decreasing the spectrum of the immune response.

5.4 Comparison of the Generated Epitopes with the Ones Identified By Other Investigators.

The amino acid epitopes generated from the three regions were back-translated to nucleotide sequences using back-translation software called Entelechon gene synthesis software available on the web under the URL http://www.entelechon.com/index.php?id=toolsand backtranslationandlang=eng and aligned using the BLAST tool to establish where they lie in the HIV genome.

Comparison of the epitopes generated in this study with the ones identified by other investigators shows that the epitopes come from almost similar locations. The epitopes from *gp 120 (*p24 region) were generated from location between 5934 and 7315 base pairs of the HIV-1 genome. From the *gag region* the epitopes were generated from the regions 1053 and 1103 base pairs of the HIV-1 genome. The epitopes from the *gp 41* region were identified from regions 4776 and 7824 base pairs of the HIV-1 genome.

Epitopes that have been previously described by other investigators were selected and their location on the HIV-1 genome determined (Table 4.13-4.15). Most of these epitopes identified by other investigators have been defined as the best CTL epitopes. These epitopes were obtained from *gp120* between 5731 and 6805 base pairs of the HIV genome. Epitopes from *gag* region lie between 714 and 6606 base pairs of the HIV genome. While those from *gp41* lie between 5494 and 8410 base pairs of the HIV genome.

In the amino acid alignment in this study there were gaps (no alignment of the sequences) where no epitopes were identified. These are regions probably with variant amino acids and hence with very little or no conserved amino acid sequences. For example, the *gag* region had gaps between 1103 and 1126, 1274 and 1310, 1181 and 1204, 1385 and 1416, and between 1442 and 1630 base pairs. *Gp 120* had gaps between 5934 and 5955, 6198 and 6236, 6176 and 6597, 6538 and 6563, 6600 and 6669, 6749 and 6864, 7122 and 7315 base pairs. There were also gaps in *gp 41* region in the following regions, between 5000 and 6589, 6854 and 6890, 6908 and 7270, 7318 and 7390, 7424 and 7576, 7620 and 7800 base pairs.

Although the epitopes generated in this study were obtained from almost similar location (*gp 120, gag* and *gp 41*) with the ones generated by other investigators, the epitopes are not identical or similar. This shows that the HIV virus is highly variable and therefore epitopes generated are different from the ones identified by other investigators. Secondly, and more importantly, some of the epitopes generated in this study were from slightly different locations in the HIV-1 genome, different from those identified by other investigators. Where there were gaps, then most likely these regions have not been studied or they were not immunogenic as shown by the SYFPEITHI program.

5.5 Designing a Super Epitope (Cross- Clade Multi Epitope Candidate Vaccine)

Most of the vaccines that have been previously developed have not been able to achieve the desired efficacy because they were not immunogenic enough. Others have not been cross-reactive to protect from other clades because of variability of the virus due to recombination. Most of the vaccines that have undergone clinical trials were mainly from subtype B clade that is not prevalent in Africa. Most of the vaccines are not home grown hence could not protect against the circulating subtypes. Therefore, because of the failures of previous vaccines in the world, use of the conserved cross-clade epitopes is hoped to be able to achieve a higher immune protection. A vaccine can be 1000 amino acid long. Because *gag* region has an extremely high level of conservation it would be vital to construct a vaccine including this region. To overcome or contend with the diversity of HIV, it would be advisable to include epitopes from the different regions of the HIV genome to enhance cross-clade immunity.

Epitope-driven vaccine concept is an attractive approach that is being successfully pursued in a number of laboratories (Nardin *et al*., 2001). These discoveries suggest that epitope–based vaccines containing a mixture of HIV-1 epitopes that are highly conserved across clades and strains of HIV-1 are feasible. This study proposes to develop a small DNA vaccine that is 300 amino acid long that contains epitopes from the three conserved regions of HIV-1 (*gag, env gp120* and *env gp 41*) that is supposed to stimulate CTL responses. CTL responses have been previously been shown to protect against HIV-1 clades.

Immunogenicity of DNA vaccines in Humans has so far not matched the expectation raised by the pre clinical experiments. To augment the immune response elicited by the multi-epitope vaccine developed in this study, immunologic adjuvants could be used that can increase the type, strength and durability of immune responses evoked. For example use of Alum increases the strength of antibody response generated by a vaccine antigen. The immunogenicity can also be improved by use of cytidine-phosphate guanosine (CpG) – motifs which are recognised by pattern recognition receptors in macrophages and dendritic cells, and also use of other adjuvants like IL-2 gene that have been shown to effectively boost the immune response. The proposed 300 amino acid DNA will also contain a strong promoter and inserted into a virus vector (CMV, Adenovirus or smallpox virus). Therefore it is hoped that this vaccine construct if successful will not only protect against circulating subtypes in Northern Kenya but also other subtypes circulating in the other parts of Kenya.

CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The HIV-1 subtype A1 is the most predominant in turkana and Mandera while subtype C is dominant in Moyale. Northern Kenya borders a number of countries, hence from this study, it can be speculated that the distribution of HIV-1 subtypes in this region may have been influenced by cross-border movement. The results also indicate the importance of carrying out surveillance studies of HIV-1 in all regions of Kenya so as to get a clear picture of subtypes in circulation. Strain surveillance is therefore an important aspect in diagnosis and vaccine development.

A total of 99 conserved epitopes from gag region (includes 38, 39, and 22 for subtypes A1, C and Drespectively) were identified, of which 32 had a score ≥ 20 and hence considered to be immunogenic. For gp 120 region a total of 286 conserved epitopes were identified, 74 of which were considered immunogenic. While for gp 41 a total of 194 conserved (including 66, 80 and 48 for subtype D, A1 and C respectively) were identified, 41 of which were considered to be immunogenic. These Conserved epitopes identified in this study may add to the efforts in the development of a multi- epitope HIV-1 vaccine. The study confirmed that *gag* region is relatively conserved in the different clades as compared to *gp41* or *gp120.*

Conserved epitope from the three regions (gag, gp120 and gp41) identified were mostly restricted to HLA –A*03, HLA –A*026, HLA –A*08 and HLA –A*0201. Therefore epitopes to be used for vaccine design should be restricted to HLA halotypes associated with disease protection from HIV infection to prevent response bias that would result from the difference in HLA genotypes. Comparison of the epitopes generated in this study with the ones identified by other investigators indicated that they were from a similar location of the HIV genome, although the epitopes are different.

Super- epitope (multi-epitope) designed in this study has a potential of making an efficacious vaccine in Kenya. Therefore the highly conserved epitopes across the clades identified from Northern Kenya can be pooled together with other epitopes from other parts of Kenya to make a vaccine that can be used in the country. Designing a multi-clade vaccine using conserved epitopes may provide a solution to the problem of HIV-1 subtype diversity. Data from this study adds to the efforts in the multi-epitope vaccine development since the approach is feasible.

6.2 Recomendations

Owing to the great antigenic diversity of HIV it is important to carry out strain surveillance studies in Northern Kenya and the rest of the country in order to get a clear picture of subtypes circulating. Therefore potential vaccines should be tailored not only based on the local viral sequence diversity but also on the local HLA allele distribution

The efficacy of the proposed construct can be evaluated *in vitro* for both MHC binding capability and T cell responses. MHC binding can be evaluated using T2 cell binding assay (Ljungren *et al.*, 1990) while T cell responses to the peptide can be measured in standard gamma-interferon release ELISpot assay (Lieberman *et al.*, 1997).

REFERENCES

Abebe, A.; Lukashov V.; Vladimir, V.; Pollakis, G.; Kliphuis, A.; Fontanet, A.; Goudsmit, J. and Jaap, A. (2001). Timing of the HIV-1 subtype C epidemic in Ethiopia based on the early virus strains and subsequent vitus diversification.*AIDS :12:1555-61.*

Abebe, A.; Lukashov V.; Vladimir, V.; Pollakis, G.; Kliphuis, A.; Fontanet, A.; Goudsmit, J. and Tobias, F. (1997). Timing of the HIV-1 subtype C epidemic in Ethiopia based on early virus strains and subsequent virus diversification. *AIDS; 15:1555-61.*

Abimiku, A.; Franchini, G.; Tartaglia, G.; Aldrich, J.; Myagkikh, K.; Markham, M.; Chong, P.; Klein, M.; Kieny, M.; Paoletti, E.; Abimiku, A. and Franchini, G (1993). HIV-1 recombinant poxvirus vaccine induces cross-protection against HIV-2 challenge in rhesus macaques. *Nat Med; 1: 321-9.*

Ada, G.; Leung, K. and Erti H. (1981). An analysis of effector T cell generation and function in mice exposed to influenza A or Sendai viruses. *Immunol Rev: 58:5–24.*

Agwale, S.; Robbins, K.; Odama, L.; Saekhou, A.; Zeh, C.; Edubio, A.; Njoku, O.; Sani-Gwarzo, N.; Gboun, M.; Gao, F.; Reitz, M.; Hone, D.; Folks, T.; Pieniazek, I.; Wambebe, C. and Kalish, M. (2002). Development of an *env* gp41-Based Heteroduplex Mobility Assay for Rapid Human Immunodeficiency Virus Type 1 Subtyping . *J. Clinic. Microbiol.; 39: 2110–14.*

Allen, T.; Herndon, J. and Wyatt, L. (2000). Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. *Nature: 407: 386–390.*

Altschul, S.; Madden, T.;_Schaffer, A.; Zhang, J.; Zhang, Z.; Miller, W. and Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res: 25:3389-402.*

Altman, J.; Moss, P.; Goulder, P.; Barouch, D.; McHeyzer-Williams, M.; Bell, J.; McMichael, A. and Davis, M. (1996). Phenotypic analysis of antigen-specific T lymphocytes. *Science; 274:94-6.*

Altuvia, Y. and Margalit, H. (2000). A structure-based approach for prediction of MHC-binding peptides. *Methods; 34:454-9.*

Amara, R.; Villinger, F.; Altman, J.; Lydy, S.; O'Neil, S.; Staprans, S.; Montefiori, D, Xu, Y.; Herndon, J. and Wyatt, L **(2001).** Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine. *Science: 292:69-74.*

Amara, R.; Villinger, F.; Altman, J.; Lydy, S.; O'Neil, S.; Staprans, S.; Montefiori, D, Xu, Y.; Herndon, J. and Wyatt, L **(2002).** Different patterns of immune responses but similar control of of a simian- human immunodeficiency virus 89.6p mucosal challenge by modified vaccinia virus Ankara (MVA) and DNA/MVA vaccines. *J. Virol; 76: 7625-31.*

An, L. and Whitton, J. (1997). A multivalent minigene vaccine, containing B-cell, cytotoxic T-lymphocyte, and Th epitopes from several microbes, induces appropriate responses in vivo and confers protection against more than one pathogen.*J Virol;71:2292-302.*

Baba, T.; Liska, V.; Hofmann-Lehmann, R.; Vlasak, J.; Xu, W.; Ayehunie, S.; Cavacini, L.; Posner, M.; Katinger, H.; Stiegler, G.; Bernacky, B.; Rizvi, T.; Schmidt, R.; Hill, L.; Keeling, M. ; Lu, Y.; Wright, J.; Chou, T. and Ruprecht, R.; (2000). Human neutralizing monoclonal antibodies of the IgG1 subtype protect against mucosal simian-human immunodeficiency virus infection. *Nat Med: 6:200–6.*

Baba, T.; Jeong, Y. and Penninck, D. (1995). Pathogenicity of live, attenuated SIV after mucosal infection of neonatal macaques. *Science: 267:1820–25.*

Barnett, S.; Lu, S.; Srivastava, I.; Cherpelis, S.; Gettie, A.; Blanchard, J.; Wang, S.; Mboudjeka, I.; Leung, L.; Lian, Y.; Fong, A.; Buckner, C.; Ly, A.; Hilt, S.; Ulmer, J.; Wild, C.; Mascola, J.; Stamatatos, L.; Barnett, S. and Lu, S. (2001). The ability of an oligomeric human immunodeficiency virus type 1 (HIV-1) envelope antigen to elicit neutralizing antibodies against primary HIV-1 isolates is improved following partial deletion of the second hypervariable region. *J Virol; 75: 5526-40.*

Barney, S. (2000). Clinical trials of HIV vaccines. Vanderbilt University School of Medicine, Nashville, TN 37232–2582 Bethesda, MD20892–300.

Barre-Sinoussi, F.; Chermann, J.; Rey, F.; Nugeyre, M.; Chamaret, S. and Gruest, J. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science: 220: 868-71.*

Barouch, D.; Fu, T.; Montefiori, D.; Lewis, M.; Shiver, J. and Letvin, N. (2001). Vaccine-elicited immune responses prevent clinical AIDS in SHIV(89.6P)-infected rhesus monkeys.*Immunol Lett; 79:57-61.*

Barouch, D.; Santra, S.; Schmitz, J.; Kuroda, M.; Tong-Ming, F.; Wagner, W.; Bilska, M.; Craiu, A.; Zheng, X.; Krivulka, G.; Beaudry, K.; Lifton, M.; Nickerson, C.; Trigona, W.; Punt, K.; Freed, D.; An, L, Dubey, S.; Casimiro, D.; Simon, A.; Davies, .; Chastain, M.; Strom, T.; Gelman, R.; Montefiori, D.; Lewis, M.; Emini, E.; Shiver, J. and Letvin, N. (2000). Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination. *Science :* 290:486–92.

Barouch, D.; Santra, S.; Steenbeke, T.; Zheng, X.; Perry, H.; Davies, M.; Freed, D.; Craiu, A.; Strom, T.; Shiver, J. and Letvin NL.(1998). Augmentation and suppression of immune responses to an HIV-1 DNA vaccine by plasmid cytokine/Ig administration. *J Immunol;161:1875-82.*

Bebenek, K.; Abbots, J.; Roberts, J.; Wilson, S. and Kunkel, T. (1989). Specificity and mechanism of error –prone replication by human immunodeficiency virus-1 reverse transcriptase. *Biol. Chem: 264:1257-61.*

Binley, J.; Ditzel, H.; Barbas, C.; Sullivan, N.; Sodroski, J.; and Burton D. (2000). Human antibody responses to HIV type 1 glycoprotein 41 cloned in phage display libraries suggest three major epitopes are recognized and give evidence for conserved antibody motifs in antigen binding. *AIDS Res Hum Retrovirol: 12:911-24.*

Boulerice, F.; Li, X.; Lvovich, A. and Wainberg, M. (1991). Recovery of infection human immunodeficiency virus type 1 after fusion of defectively infected clones of U-937 cells. *J. Virol: 65: 5589-5592.*

Borrow, P.; Lewicki, H.; Wei, X.; Horwitz, M.; Peffer, N.; Meyers, H.; Nelson, J.; Gairin, J.; Hahn, B.; and Shaw, G. (1997). Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med: 3:205-11.*

Buve, A.; Carael, M.; Hayes, R. and Robinson, N. (1995). Variations in HIV prevalence between urban areas in sub-Saharan Africa: do we understand them? *AIDS: 9:S103-9.*

Burton, D.; Williamson, R. and Parren P. (2000). Antibody and virus: binding and neutralization. *Virology: 270:1–3.*

Burton, D. (2002). Antibodies, viruses and vaccines. *Nature Rev. Immunol: 2:706–13.*

Burton, D. (1997). A vaccine for HIV type 1: The antibody perspective. *Proc Natl Acad Sci USA: 94:10018-23.*

Blackard, J.; Cohen, D.; and Mayer, K. (2002). Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences. *Clin.Infect.Dis: 4: 1108-14.*

Bliss, J.; Van Cleave, V.; Murray, K.; Wiencis, A.; Ketchum, M.; Maylor, R.; Haire, T.; Resmini, C.; Abbas, A. and Wolf, S.(1996). IL-12, as an adjuvant, promotes a T helper 1 cell, but does not suppress a T helper 2 cell recall response. *J Immunol; 156:887-94.*

Brander, C. (2004). HLA and HIV: Implications for HIV Vaccine Design. Partners AIDS Research Center. Massachusetts General Hospital. Harvard Medical School, Boston, USA.

Brander, C., and P. Goulder. (1999). Recent advances in HIV-1 CTL epitope charactrization. In HIV Molecular immunology database. Theoretical Biology and Biophysics, Los Alamos, NM, USA, pp1-1.

Broliden, P.; von Gegerfelt, A.; Clapham. P.; Rosen, J.; Fenyo, E.; Wahren, B. and Broliden, K. (1992). Identification of human neutralization-inducing regions of the human immunodeficiency virus type 1 envelope glycoproteins. *Proc Natl Acad Sci U S A: 89:461-5.*

Carr, J.; Salminen, M.; Albert, J.; Sanders-Buell, E.; Birx, D. and McCutchan, F. (1998). Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology: 247:22-31.*

Carr, J.; Laukkanen, T.; Salminen, M.; Albert, J.; Alaeus, A.; Kim, B.; Sanders-Buell, E.; Birx, D. and McCutchan, F. (1999). Characterization of subtype A HIV-1 from Africa by full genome sequencing. *AIDS; 13:1819-26.*

Carr, J. ; Salminen, M.; Koch, C.; Gotte, D.; Artenstein, A.; Hegerich, P.; St. Louis, D.; Burke, D and McCutchan, F. (1996). Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol; 70:5935–5943.*

Carrington, M., G. Neslon, M. P. Martin, T. Kissner, D. Vlahov, J. J. Goedert, R. Kaslow, S. Buchbinder, K. Hoots, and S. J. O'Brien. (1999). HLA and HIV-1: heterozygote advantage and B*35-CW*04 disadvantage. *Science 283:1748.*

Cavacini, L.; Samore, M.; Gambertoglio, J.; Jackson, B.; Duval, M.; Wisnewski, A.; Hammer, S.; Koziel, C.; Trapnell, C. and Posner, M. (1998). Phase I study of a human monoclonal antibody directed against the CD4-binding site of HIV type 1 glycoprotein 120. *AIDS Res Hum Retrovirol: 14:545-50.*

Cease, K. and Berzofsky, J. (1994). Toward a vaccine for AIDS: the emergence of immunobiology-based vaccine development. *Annu Rev Immunol; 12: 923-89.*

Chakrabarti, B.; Maitra, R. and Ma, X. (1996). A candidate live inactivatable attenuated vaccine for AIDS. *Proc Natl Acad Sci USA:93:9810-15.*

Chaturvedi, P.; Yu, Q.; Southwood, S.; Sette, A. and Singh, B. (1996). Peptide analogs with different affinities for MHC alter the cytokine profile of T helper cells. *Int. Immunol:* 8:745–55.

Check, E. (2003). Vaccines: back to 'plan A'. *Nature: 912–914.*

Chomczynki, P.; Mackey, K.; Drews, R. and Wilfinger, W. (1997). DNAzol reagent for the rapid isolation of genomic DNA.*Bio Technique; 22:550-553.*

Choudhury, S.; Montano, M.; Womack, C.; Blackard, J.; Maniar, J.; Saple, G.; Tripathy, S.; Sahni, S.; Shah, G.; Babu, P. and Essex, M. (2000). Increased promoter diversity reveals a complex phylogeny of human immunodeficiency virus type 1 subtype C in India. *J. Hum. Virol; 3:35–43.*

Clerici, M.; Levin, J.; Kessler, H.; Harris, A.; Berzofsky, J.; Landay, A. and Shearer, G.; (1994). HIV-specific T-helper activity in seronegative health care workers exposed to contaminated blood. *JAMA: 271:42–6.*

Constant, S.; Sant'Angelo, D.; Pasqualini, T.; , T.; Levin, D.; Flavell, R. and Bottomly, K. (1995). Peptide and protein antigens require distinct antigen presenting cell subsets for the priming of CD4$^+$ T cells. *J. Immunol: 154:4915–23.*

Cooper, N. (1994). Interaction of the Complement System with Microorganisms.*Landes, 133–50*

Cotropia, J.; Ugen, K.; Kliks, S.; Broliden, K.; Broliden, P.; Hoxie, J.; Srikantan, V.; Williams, W. and Weiner, D. (1996). A human monoclonal antibody to HIV-1 gp41 with neutralizing activity against diverse laboratory isolates. *J Acquir. Immune. Defic. Syndr.Hum. Retroviro: 12:221-32.*

Cullen, B. (1998). HIV-1 auxiliary proteins: making connections in a dying cell. *Cell: 29: 93:685-92.*

Davis, N.; Caley, I.; Brown. K.; Betts. M.; Irlbeck. D.; McGrath. K.; Connell, M.; Montefiori, D.; Frelinger, J. and Swanstrom, R. (2000). Vaccination of macaques against pathogenic simian immunodeficiency virus with Venezuelan equine encephalitis virus replicon particles. *J Virol: 74:371-378.*

Daniel, M.; Kirchhoff, D.; Czajak, C.; Sehgal, P and Desrosiers. R. (1992). Protective effects of a live attenuated SIV vaccine with a deletion in the nef gene. *Science; 258:1938–1941.*

Damond, F.; Apetrei, C.; Robertson, D.; Souquiere, S.; Lepretre, A. and Matheron S. (2001). Variability of human immunodeficiency virus type 2 (HIV-2) infecting patients living in France. *Virol: 280: 19-30.*

Deacon, N.; Tsykin, A, Solomon, A.; Smith, K.; Ludford-Menting, M.; Hooker, D.; McPhee, D.; Greenway, A.; Ellett, A.; Chatfield, C.; Lawson, V.; Crowe S, Maerz, A, Sonza S, Learmont J, Sullivan JS, Cunningham A, Dwyer D, Dowton D, Mills J. (1995). Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients. *Science: 10: 270:988-91.*

De Cock, K.; Adjorlolo, G.; Ekpini, E.; Sibailly, T.; Kouadio, J.; Maran, M.; Brattegaard, K.; Vetter, K.; Doorly, R. and Gayle, H. (1994). Epidemiology and transmission of HIV-2. Why there is no HIV-2 pandemic. *J.Ameri.Med.Ass: 270:2083-6.*

De Groot, S.; Sbai, H.; Frost, J.; Saint-Aubin, C.; Chinai, N.; Martin, W.; Bosma, A.; Skowron, G. and Mayer, K. (2001). Designing HIV-1 vaccines to reflect viral diversity and the global context of HIV/AIDS. *AIDScience: 1:210-295.*

De Pasquale, M.; Leigh Brown, A.; Uvin, S.; Allega-Ingersoll, J.; Caliendo, A.; Sutton, L.; Donahue, S. and D'Aquaila, R. (2003). Differences in HIV-1 pol sequence from female genital tract and blood during antiretroviral therapy. *J.Acquied Immune Defic.Synd: 34: 37-44.*

Desrosiers, R. C. (2004). Prospects for an AIDS vaccine. *Nat Med 10(3): 221-3.*

Dierich, M.; Marschang, P.; Larcher, C. (1992). HIV and human complement: molecular mechanisms and biological consequences. In Progress in Immunology. *VIII ed. J Gergely: 691–98.*

Dimmock, N. (1993). Neutralization of animal viruses. *Curr. Top. Microbiol. Immunol: 83:1–149.*

Donnes P, Elofsson A. (2002). Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics: 11; 3:25.*

Domingo, E. (1998). Quasispecies and the implication for virus persistence and escape. *Clin. Diagn. Virol: 10: 97-101.*

Dowbenko, D.; Nakamura, G.; Fennie, C.; Shimasaki, C.; Riddle, L.; Harris, R.; Gregory, T. and Lasky, L. (1998). Epitope mapping of the human immunodeficiency virus type 1 gp120 with monoclonal antibodies. *J Virol: 62:4703-11.*

Dowling, W.; Kim, B.; Mason, C. and Barrett, W. (2002). Forty-one full length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A containing recobinants. *AIDS: 16:1809-1820.*

Dyer, W.; Ogg, G.; Menoitie, M.; Jin, X.; Geczy, A.; Rowland-Jones, S.; McMichael, A.; Nixon, D. and Sullivan. (1999). Strong human immunodeficiency virus (HIV)-specific cytotoxic T-lymphocytes activity in Sydney Blood Bank Cohort patients infected with nef-defective HIV-type 1. *J Virol: 73:436–43.*

Elliott, T.; Driscoll, P.; Smith, M. and McMichael, A. (1994). Peptide epitope selection by class I MHC molecules. *Curr. Biol: 3:854M-P2.*

Emini, A.;Schleif, W.;Nunberg, J.;Conley,A.;Eda,Y.; Tokiyoshi,S.Putney, D.Matsushita S, Cobb KE, Jett CM (1992). Prevention of HIV-1 infection in chimpanzees by gp120 V3 domain-specific monoclonal antibody. *Nature: 355: 728–730.*

Emini, E.; (1995). Resistance to anti-human immunodeficiency virus therapeutic agents. *Adv Exp Med Biol: 390:187-95.*

Evans, D.; O'Connor, D.; Seth, A. and Ourmanov, I. (1999). Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. *Nat Med; 5: 1270-6.*

Feng, T.; Ni, A.; Yang, G.; Galvin, S.; Hoffman, I. and Cohen, M. (2003). Distribution of the CCR5 gene 32-base pair deletion and CCr5 expression in Chinese minorities. *J Acquired Immune Defic Syndr: 32: 131-4.*

Feng, Y.; Broder, C.; Kennedy, P.; and Berger, E. (1996). HIV-1 entry cofactor: functional cDNA cloning of a seven transmembrane, G protein-coupled receptor. *Science: 272: 872–877.*

Fisher, A.; Ensoli, B.; Looney, D.; Rose, A.; Gallo, R.; Saag, C.;Shaw, S.; Hahn, G. and Wong-Staal, F. (1988). Biologically diverse molecular variants within a single HIV-1 isolate. *Nature; 334: 444-7.*

Fowke, K.; Nagelkerke, N. and Kimani, J. (1996). Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet: 348:1347–51.*

Fletcher, C.; Acosta, E.; Cheng, H.; Haubrich, R.; Fischl, M.; Raasch, R.; Mills, C.; Hu, X.; Katzenstein, D.; Remmel, R. and Gulick, R. (2000). Competing drug-drug interactions among multidrug antiretroviral regimens used in the treatment of HIV-infected subjects: *AIDS: 14:2495-501.*

Frahm, N., P. Goulder, and C. Brander. (2002). Total assessment of HIV specific CTL responses: Epitope clustering, processing preferences and the impact of HIV sequence heterogeneity. In HIV Molecular immunology database, Los Alamos National Laboratory: Theoretical Biology and Biophysics, Los Alamos, NM, USA.

Francis, D. (2001). AIDSVAX is safe, but efficacy question remains. VaxGen founder discusses quest for AIDS vaccine. *AIDS; 16: 141-2.*

Francis, D.; Gregory, T.; McElrath, M.; Belshe, R.; Gorse, G.; Migasena, S.; Kitayaporn, D.; Pitisuttitham, P.; Matthews, T.; Schwartz, D.; Berman, P.; Francis, D.; Gregory, T. (1998). Advancing AIDSVAX to phase 3. Safety, immunogenicity, and plans for phase 3. *AIDS Res Hum Retrovir; 14l 3: 325-31.*

Fred, G.; Albert-Edward, C.; Bronson, D.; Fitzgerald, J.; Anderson (1995). Circular structures in retroviral and cellular Genomes. *The Journal of biological Chemistry: 70:23570-23581.*

Gao, F.; Weaver, E.; Lu, Z.; Li, Y.; Liao, H.; Ma, B.; Alam, S.; Scearce, R.; Sutherland, L.; Yu, J.; Decker, J.; Shaw, G.; Montefiori, D.; Korber, B.; Hahn, B and Haynes, B.(2005). Antigenicity and immunogenicity of a synthetic human immunodeficiency virus type 1 group m consensus envelope glycoprotein. *J Virol: 79:1154-63.*

Gao, F.; Robertson, D.; Morrison, S.; Hui, H.; Craig, S.; Decker, J.; Fultz, P.; Girard, M.; Shaw, G.; Hahn, B.; Sharp, P. (1996). The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol: 70: 7013-29.*

Gallo, R.; Sarin P.; Gelmann E.; Robert-Guroff M.; Richardson E. and Kalyanaraman, V. (1983). Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science: 220: 865-7.*

Gaschen, B.; Taylor, J.; Yusim, K.; Foley, B.; Gao, F.; Lang, D.; Novitsky, V.; Haynes, B.; Beatrice, H.; Bhattacharya, T.; Korber, B. (2002). Diversity consideration in HIV –1 vaccine selection. *Science: 2: 296-299.*

Goudsmit, J. (1988). Immunodominant B-cell epitopes of the HIV-1 envelope recognized by infected and immunized hosts. *AIDS: S41–S45.*

Goulder, P.; Jeena, P.; Tudor-Williams, G. and Burchett, S. (2001). Paediatric HIV infection: correlates of protective immunity and global perspectives in prevention and management. *Br Med Bull: 58:89-108.*

Goulder, P.; Phillips, R.; Colbert, R.; McAdam, S.; Ogg, G.; Nowak, M.; Giangrande, P.; Luzzi, G.; Morgan, B.; Edward, A.; McMichael, A. and Rowland-Jones, S. (1997). Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat Med: 3:212-7*

Golovkina, T. V.; Jaffe, A. B. and Ross, S. R. (1994). Coexpression of exogenous and endogenous mouse mammary tumor virus RNA in vivo results in viral recombination and broadens the virus host range. *J.virol: 8: 5019-5026.*

Gorny, M.; Xu, J.; Gianakakos, V.; Karwowska S, Williams, C.; Sheppard, H.; Hanson, C.; Zolla-Pazner, S. (1991). Production of site-selected neutralizing human monoclonal antibodies against the third variable domain of the human immunodeficiency virus type 1 envelope glycoproteins. *Proc Natl Acad Sci U S A: 15; 88:3238-42.*

Graham, B.; Bunton, L.; Wright, P. and Karzon, D. (1991). The role of T cell subsets in the pathogenesis of primary infection and reinfection with respiratory syncytial virus in mice. *J Clin Invest: 88:1026–33.*

Hanke, T. and McMichael, A. (2000). Design and construction of an experimental HIV-1vaccine for the year –2000 clinical trials in Kenya. *Nat. Med: 6:951-955.*

Haigwood, N.; Watson, A.; Sutton, W.; McClure, J.; Lewis, A.; Ranchalis, J.; Travis, B.; Voss, G.; Letvin, N.; Hu, S.; Hirsch, V. and Johnson, P. ***(1996).*** Passive immune globulin therapy in the SIV/macaque model: early intervention can alter disease profile. *Immunol. Lett: 5: 107–14.*

Haynes, B.; Pantaleo, G. and Fauci, A. (1996). Towards an understanding of the correlate of protective immunity to HIV infection. *Science: 271: 324-328.*

Hanke, T.; McMichael, A.; Mwau, M.; Wee, E.; Ceberej, I.; Patel, S.; Sutton, J.; Tomlinson, M. and Samuel, R. (2002). Development of a DNA-MVA/HIVA vaccine for Kenya. *Vaccine: 20:1995-8.*

Hazenberg, M.; Hamann, D.; Schuitemaker, H. and Miedema, F. (2000). T cell depletion in HIV-1 infection: how CD4+ T cells go out of stock. *Nat Immuno: 1:285-9.*

Heeney, J.; Akeblom, L. and Hahn, B. (1995). Vaccines and immunology: elucidating immunity to HIV-1 and current prospects for AIDS vaccine development. *AIDS: 14: 125-7.*

Heeney, J.; Teeuwsen, M.; van Gils, W.; Bogers, C.; Morghen, A.; Radaelli, S.; Barnett, B.; Morein, L.; Akerblom, Y.; Wang, T.; Lehner, W. and Davis, D. (1998).. Beta-chemokines and neutralizing antibody titers correlate with sterilizing immunity generated in HIV-1 vaccinated macaques. *Proc. Natl. Acad. Sci: 95:10803–08.*

Heyndrickx, L.; Janssens, W.; Ndumbe, P.; Vereecken, K.; Coppens, S.; De Houwer, K.; and Van der Groen, G. (2000). HIV-1 genetic variability in Cameroon. *AIDS: 14:1862-4.*

Hess, J. (2005). Uncross-matched red blood cells save lives. *Hosp Med. Feb: 66:95-6.*

Herberman, R.; Reynolds, C. and Ortaldo, J. (1986). Mechanism of cytotoxicity by natural killer (NK) cells. *Annu. Rev. Immunol: 4:651–80.*

Heribert, S. and Manfred, P. (1997). Role of complement in HIV infection. *Annual Review of Immunology: 15: 649-674.*

Hirsch, V.; Sharkey, M.; Brown, C.; Brichacek, B.; Goldstein, S.; Wakefield, J.; Byrum, R.; Elkins, W.; Hahn, B.; Lifson, J. and Stevenson, M. (1998). Vpx is required for dissemination and pathogenesis of SIV(SM) PBj: evidence of macrophage-dependent viral amplification. *Nat Med: 4:1401-8.*

Hilleman, M. (2001). Overview of the pathogenesis, prophylaxis and therapeutics of viral hepatitis B, with focus on reduction to practical applications. *Vaccine: 19: 1837–48.*

Hirsch, V.; Fuerst, T. and Sutter G, (1996). Patterns of viral replication correlate with outcome in simian immunodeficiency virus (SIV)-infected macaques: effect of prior immunization with a trivalent SIV vaccine in modified Vaccinia virus Ankara. *J Virol: 70:3741-3752.*

Ho, D.; Fung, M.; Cao, Y.; Li, X.; Sun, C.; Chang, T. and Sun, N. (1991). Another discontinuous epitope on glycoprotein gp120 that is important in human immunodeficiency virus type 1 neutralization is identified by a monoclonal antibody. *Proc Natl Acad Sci U S A: 15: 88:8949-52.*

Holm-Hansen, C.; Stern, B.; Rustad, S.; Shao, J. and Asjo B. (2000). V3 sequence analysis and biological characterization of HIV-1 isolates from asymptomatic and early symptomatic Tanzanian individuals. *APMIS: 108:608-16.*

Hosein S. (1995). HIV/AIDS Theory: HIV and cytokines. *Community AIDS treatment information: 55:55-56.*

Hu, P.; Hultin, L.; Hultin, P.; Hausner, M.; Hirji, K.; Jewett, A.; Bonavida, B.; Detels, R. and Giorgi, J. (1995). Natural killer cell immunodeficiency in HIV disease is manifest by profoundly decreased numbers of $CD16^{+}CD56^{+}$ cells and expansion of a population of $CD16^{dim}CD56^{-}$ cells with low lytic activity. *J. Acq. Immun. Deficiency Syndrome Hum. Retrovirol: 10:331–40.*

Hu, P.; Downing, R.; Hultin, L.; Hultin, P.; Hausner, M.; Hirji, K.; Jewett, A.; Bonavida, B.; Detels, R. and Giorgi, J. (2000). Predominance of HIV-1 subtype A and D infections in Uganda. *Emerg Infect Dis: 6: 609-615.*

Jankovic, D.; Caspar, P.; Zweig, M.; Garcia-Moll, M.; Showalter, S.; Vogel, F. and Sher, A. (1997). Adsorption to aluminum hydroxide promotes the activity of IL-12 as an adjuvant for antibody as well as type 1 cytokine responses to HIV-1 gp120. *J Immunol;159:2409-17.*

Jason, T.; Blackard, D. and Mayer, K. (2002). Human Immunodeficiency Virus Superinfection and Recombination: current status of knowledge and potential clinical consequences. *Clinical Infectious Diseases: 34:1108-14.*

Janssens, W.; Buve, A.; Motte, J.; Piot, P. and Ven der Groen, G. (1992). Near full-length genome analysis of HIV-1 CRF AG and AC- subrype G recombinants.AIDS Res Hum Rtrovirus: 10: 277-79.

Janssens, W.; Buve, A. and Nkengasong, J. (1997). The puzzle of HIV-1 subtypes in Africa. *AIDS: 11: 705–712.*

Janssens, W.; Heyndrickx, L.; Fransen, K.; Temmerman, M.; Leonaers, A.; Ivens, T.; Motte, J.; Piot, P. and Ven der Groen, G. (1994). Genetic variability of HIV type 1 in Kenya. *AIDS research and human retroviruses: 10:1577-79.*

Jawetz, L. (2000). Medical Microbiology and Immunology: Examination and Board Review. Sixth Ed: Lange/McGraw-Hill.

Johnson, R.; Lifson, J.; Czajak, S.; Cole, K.; Manson, K.; Glickman, R.; Yang, J.; Montefiori, D.; Montelaro, R.; Wyand, M. and Desrosiers, R.(1999). Highly attenuated vaccine strains of simian immunodeficiency virus protect against vaginal challenge: inverse relationship of degree of protection with level of attenuation. *J Virol: 73:4952-61.*

Jorgensen, J.; Reay, P.; Ehrich, E. and Davis, M. (1992). Molecular components of T-cell recognition. *Annu. Rev. Immunol: 10:835–73.*

Kahn, J.; Sinangil, F.; Baenziger, J.; Lockey, T.; Howlett, N.; Srivinas, R.; Rencher, S.; Freiden, P.; Doherty, P. (1994). Clinical and immunologic responses to human immunodeficiency virus (HIV) type 1(SF2) gp120 subunit vaccine combined with MF59 adjuvant with or without muramyl tripeptide dipalmitoyl phosphatidylethanolamine in non-HIV-infected human volunteers. *J Infect Dis; 170:1288-91.*

Kandathil, S.; Ramalingam, R.; Kannangai, S.; and Sridharan, D. (2005). Molecular epidemiology of HIV. *Indian J Med Res: 121: 333-344.*

Kanki, P.; Travers, K.; Mboup, S.; Hsieh, C.; Marlink, R.; Gueye-Ndiaye, A.; Siby, T.; Thior, I.; Hernandez-Avila, M. and Sankale, J. (1994). Slower heterosexual spread of HIV-2 than HIV-1. *Lancet: 343:943-6.*

Kalams S.; Goulder P.; Shea A.; Jones N.; Trocha A.; Ogg G. and Walker B. (1999). Levels of human immunodeficiency virus type 1-specific cytotoxic T-lymphocyte effector and memory responses decline after suppression of viremia with highly active antiretroviral therapy. *J Virol: 73:6721-8.*

Karlin, S.; Blaisdell, B. and Brendel, V. (1990). Identification of significant sequence patterns in proteins. *Methods. Enzymol; 183:388-402.*

Karlin, S. and Altschul, S. (1990). Method for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc.Natl.Acad.sci.USA; 87:2264-2268.*

Karlin, S. and Altschul, S. (1993). Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc. Natl. Acad. Sci. U S A; 90:5873-7.*

Keefer, M.; Graham, B.; Belshe, R.; Freiden, P.; Doherty, P. and Hurwitz, J. (1994). Studies of high doses of a human immunodeficiency virus type 1 recombinant glycoprotein 160 candidate vaccine in HIV type 1-seronegative humans. *AIDS Res Hum Retrovir; 10:1713-23.*

Kelleher, A. and Bockel, A. (2005) BUSINESS BRIEFING: European pharmarcotherapy.

Kestler, H.; Ringler, D.; Mori, K.; Panicali, D.; Sehgal, P.; and Desrosiers, R.; (1991). Importance of the nef gene for maintenance of high virus loads and for development of AIDS. *Cell: 65:651-62.*

Khamadi, S.; Ochieng, W.; Lihana, R.; Kinyua, J.; Muriuki, J.; Mwangi, J.; Lwembe, R.; Kiptoo, M.; Osman, S.; Lagat, N.; Pelle, R.; Muigai, A.; Carter, J.; Oishi, I.; Ichimura, H.; Mwaniki, D.; Okoth, F.; Mpoke, S. and Songok, E. (2005). HIV type 1 subtypes in circulation in northern Kenya. *AIDS Res Hum Retroviruses; 21:810-4.*

Kirchhoff, F.; Greenough, T.; Brettler, D.; Sullivan, J.; Desrosiers, R. (1995). Brief report: Absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1infection. *N Engl J Med: 332:228–32.*

Klenerman P. Rowland-Jones S. McAdam S. et al. (1994). Cytotoxic T cell activity antagonized by naturally occurring HIV-1 gag variants. *Nature. 369:403.*

Klenerman, P. and Zinkernagel, R. (1998). Original antigenic sin impairs cytotoxic T lymphocyte responses to viruses bearing variant epitopes. *Nature; 394:482-5.*

Kolchinsky, P.; Mirzabekov, T.; Farzan, M.; Kiprilov, E.; Cayabyab, M.; Mooney, L.; Choe, H. and Sodroski, J. (1999). Adaptation of a CCR5-using, primary human immunodeficiency virus type 1 isolate for CD4-independent replication. *J Virol: 73:8120-6.*

Koup, R.; Safrit, J.; Cao, Y. (1994). Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J Virol : 68:4650–55.*

Korber, B.; Gaschen, B.; Yusim, K.; Thakallapally, R.; Kesmir, C. and Detours, V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull: 58:19-42.*

Korber, B.; Gaschen, B.; Yusim, K.; Thakallapally, R.; Kesmir, C. and Detours, V. (2002). Evolutionary and immunological implications of contemporary HIV-1 variation. Br-Med. *Bull, 58:19-42.*

Kuiken, C.; Fley, B.; Freed, E. and Hahn, B. (2002). HIV sequence compendium. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Kuller, L.; Benveniste, C.; Tsai, E.; Clark, P.; Polacino, R.; Watanabe, J.; Overbaugh, M.; Katze W. and Morton, R. (1994). Intrarectal inoculation of macaques by the simian immunodeficiency virus, SIVmne E11S: CD4 depletion and AIDS. *J. Med. Primatol: 23:397–409.*

Landay, A. and Heeney, J. (2001). Vaccines and immunology. Recent advances in HIV/AIDS vaccine development. *AIDS: 5:S101-3.*

LeGros, G.; Ben-Sasson, S.; Seder, R.; Finkelman, F. and Paul, W. (1990). Generation of IL4 producing cells in vivo and in vitro. IL2 and IL4 are required for in vitro generation of IL4 producing cells. *J. Exp. Med: 172:921–29.*

Letvin, J. (1998) What Immunity Can Protect against HIV Infection? *Clin. Invest: 102: 1643–1644.*

Lewis, D.; Tang, D.; Adu-Oppong, A.; Schober, W. and Rodgers, J. (1994). Anergy and apoptosis in CD8+ T cells from HIV-infected persons. *J Immunol: 153:412-20*

Ljunggren, H.; Stam, N.; Rotzschke, O.; Falk, K.; Neefjes, J.; Hoglund, P.; Heemels, M.; Bastin, J et al. (1990). Empty MHC class I molecules come out in the cold. *Nature; 346: 476-80.*

Lieberman, J.; Fabry, J.; Rotzschke, O.; Falk, K.; Neefjes, J.; Hoglund, P.; Heemels, M. and Bastin, J (1997). Recognition of a small number of diverse epitopes dominates the cytotoxic T lymphocytes response to HIV type 1 in an infected individual. *AIDS Res Hum Retroviruses; 13: 383-92.*

Lipford, G.; Hoffman, M.; Zinkernagel, R. and Doherty, P. (1993). Primary in vivo responses to ovalbumin. Probing the predictive value of the Kb binding motif. *J Immunol; 150: 1212-22.*

Lu, L. and Askonas, B. (1980). Cross-reactivity for different type A influenza viruses of acloned T-killer cell line. *Nature: 288:164–5.*

Luis, C.; Afonso, C.; Scharton, T.; Vieira, L.; Wysocka, M.; Trinchieri, G. and Scott, P. (1994). The adjuvant effect of interleukin-12 in a vaccine against Leishmania major. *Science;263:235-37.*

Machuca, A.; Soriano, V.; Gutirrez, M.; Holguin, A.; Aguilera, A. and Caballero, E. (1999). Human immunodeficiency virus type 2 infection in Spain. The HIV-2 Spanish Study Group. *Intervirology: 42: 37-42.*

Mahmudul, H. and Mahbub-E-Sobhani, (2003). Different methods of HIV Vaccination, efficacy and their delivery system: A review. *J. Med. Sci: 3:1-23.*

Mascola, J.; Frankel, S. and Broliden, K. (2000). HIV-1 entry at the mucosal surface: role of antibodies in protection. *AIDS: 14: S167-74.*

Mascola, J.; Lewis, M.; Stiegler, G.; Harris, D.; VanCott, T.; Hayes, D.; Louder, M.; Brown, C.; Sapan, C.; Frankel, S.; LuY, M.; Katinger H, Birx, D. (1999). Protection of macaques against pathogenic simian/human immunodeficiency virus 89. 6PD by passive transfer of neutralizing antibodies. *J Virol: 73:4009–18.*

Mathew, H.; Barbas, C.; Parren, P.; Burton, D.; Moore, J. and Ho, D. (2002). AIDS research and Human retroviruses. *Emerg Infect Dis: 18: 1163-66.*

McCathy, C.; Sidorov, I.; Broder, C.; Robinson, J. and Parren, P. (1995). Pespectives on AIDS.HIV: between past and Future. *J. Medic. Virol: 46:18-20.*

McCune. J. (2001). The dynamics of CD4+ T-cell depletion in HIV disease. . *Nature: 410:19.*

McCutchan, F. (2000). Understanding the genetic diversity of HIV-1. *AIDS: 14: S31-S44.*

McCutchan, F. Carr, J.; Laukkanen, T.; Bednarek, M. and Coligan, J. (1999). Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. *Virology 254: 226-34.*

McMichael, A. (2006). HIV vaccine. *Annu Rev Immunol: 24:227-55.*

McMichael, A. and Hanke, T. (1999). The quest for an AIDS vaccine: is the CD8+ T-cell approach feasible? Nat Rev Immunol: 2:283–91.

McMichael, A. and Phillips, R. (1997). Escape of Human Immunodeficiency Virus from immune control. *Annu Rev Immunol: 15:271-96.*

McMichael, A. and Rowland-Jones, S. (2001). Cellular immune responses to HIV. *Nature: 410:980-7.*

Miller, C.; Marthas, J.; Torten, N.; Alexander, J.; Moore, G.; Doncel, and Hendrickx, A. (1994). Intravaginal inoculation of rhesus macaques with cell-free simian immunodeficiency virus results in persistent or transient viremia. *J. Virol:* 68: 6391–6400.

Mims, C.; Playfair, J.; Roitt, I.; Wakelin, D. and Williams, R. (1993). Medical Microbiology. 2416.

Mo, H.; Stamatatos, L.; Ip, J.; Barbas, C.; Parren, P.; Burton, D.; Moore, J. and Ho, D. (1997). Human immunodeficiency virus type 1 mutants that escape neutralization by human monoclonal antibody IgG1b12. *J Virol: 71:6869-74.*

Moulard, M.; Phogat, S.; Shu, Y.; Labrijn, A.; Xiao, X.; Binley, J.; Zhang, M.; Sidorov, I.; Broder, C.; Robinson, J.; Parren, P.; Burton, D.; Dimitrov, D. (2002).

Broadly cross-reactive HIV-1-neutralizing human monoclonal Fab selected for binding to gp120-CD4-CCR5 complexes. *Proc Natl Acad Sci U S A: 99:6913-8.*

Moutouh, L.; Corbeil, J. and Richman, D. D. (1996). Recombination leads to rapid emergence of HIV-1 Dually resistant mutants under selective drug pressure. *Proc.Natl Acad SciUSA: 93:6106-11.*

Morison, L.; Buve, A.; Zekeng, L.; Heyndrickx, L.; Musonda, R.; Kahindo, M.; Weiss, H.; Hayes, R.J.; Laga, M.; Janssens, W. and Van der Goen, G. (2001). HIV-1 subtypes and the HIV epidemics in four cities in sub-Saharan Africa. *AIDS: 15: S109-S116.*

Mokili, J.; Wade, C.; Burns, S.; Cutting, W.; Bopopi, J.; Green, S.; Peutherer, J.; Simmonds, P.; Mokili, J. and Wade, C. (1999). Genetic heterogeneity of HIV type 1 subtypes in Kimpese, rural Democratic Republic of Congo. *AIDS Res Hum Retroviruses; 15: 655-64.*

Montaner, L. and Gordon, S. (1994). Th2-mediated HIV1 virostatic state: macrophage-specific regulation in vitro. *Res. Immunol: 145:583–87.*

Montefiori, D.; Cornell, R.; Zhou, J.; Zhou, J. and Hirsch, V. (2001). Complement control proteins, CD46, CD55, and CD59, as common surface constituents of human and simian immunodeficiency viruses and possible targets for vaccine protection. *Virol; 205:82-92.*

Moss, R.; Wallace, M.; Steigbigel, R.; Morrison, S.; Giermakowska, W.; Nardo, C.; Diveley, J.; Carlo, D.; Moss, R. and Wallace, M. (2002). Predictors of HIV-specific lymphocyte proliferative immune responses induced by therapeutic vaccination. *Clin Exp Immunol; 128: 359-64.*

Mosmann, T. and Sad, S. (1996). The expanding universe of T-cell subsets: Th1, Th2 and more. *Immunol. Today: 17:138–46.*

Muster, T.; Steindl, F.; Purtscher, M.; Trkola, A.; Klima, A.; Himmler, G.; Ruker, F.; Katinger, H. (1997). A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J Virol: 67:6642-7.*

Mwau, M. and McMichael, A. (2003). A review of vaccines for HIV prevention. *J Gene Med: 5:3-10.*

Myers G. (1994). Tenth anniversary perspectives on AIDS.HIV: between past and future. *AIDS res hum retrovirus: 10:1317-24.*

Najera, R.; Delgado, E.; Peres-Alvares, L. and Thomson, M. (2002). Genetic recombination and its role in the development of the HIV-1 pandemic. Lippincott Williams and Wilkins, Inc.

Natuk, R.; Lubeck, L.; Chanda, P.; Chengalvala, M.; Wade, M.; Murthy, S.; Wilhelm, J.; Vernon, S.; Dheer, S. and Mizutani, S. (1993). Immunogenicity of recombinant human adenovirus-human immunodeficiency virus vaccines in chimpanzees. *AIDS Res Hum Retrovir; 9: 395-404.*

Nardin, E.; Oliveira, G.; Calvo-Calle, J. and Nussenzweig, R. (2001). The use of multiple antigen peptides in the analysis and induction of protective immune responses against infectious diseases. *Adv. Immunol: 60:105-49.*

Neilson, R.; John, C.; Lewis, P.; Kreisss, J.; Jackson S.; Nduati, R.; Mbori-Ngacha, D.; Panteleeff D.; Bodrug, S.; Giahetti, C.; Bott A.; Richardson, A.; Bwayo, Y.; Ndinya-Achola, J. and Overbaugh, J. (1999). Subtypes of Human Immunodeficiency Virus type 1 and disease stage among women in Nairobi, Kenya. *J. Virol: 73: 4393-4403.*

Novitsky, V.; Smith, R.; Gilbert, P.; McLane, R.; Chigwedere, P.; Williamson, C.; Ndung'u, T.; Klein, I.; Chang, S.; Peter, T.; Thior, L.; Foley, B.; Gaolekwe, S.; Rybak, N.; Gaseitsiwe,S.; Vannberg, F.;Marlink, R.; Lee, T. and Essex, M. (2002). Human Immunodeficiency Virus Type 1 Subtype C Molecular Phylogeny:Consensus Sequence for an AIDS Vaccine Design? *J. Virol;11 5435–51*

Novitsky, V.; Gaolekwe, S.; McLane, M.; Ndung'u, T.; Foley, B.; Vannberg, F.; Marlink, R. and Essex, M. (2000). HIV type 1 A/J recombinant with a pronounced pol gene mosaicism. AIDS Res. Hum. *Retrovir; 16:1015–20.*

Novitsky, V.; Montano, M.; McLane, M.; Renjifo, B.; Vannberg, F.; Foley, B.; Ndung'u, T.; Rahman, M.; Makhema, M.; Marlink, and Essex, R. (1999). Molecular cloning and phylogenetic analysis of HIV-1 subtype C: a set of 23 full-length clones from Botswana. *J. Virol; 73:4427–32.*

Ondoa, P.; Kestens, L.; Davis, D.; Vereecken, C.; Willems, B.; Fransen, K.; Vingerhoets, J.; Zissis, G.; Ten Haaft, P.; Heeney, J. and van der Groen, G. (2001). Longitudinal comparison of virus load parameters and CD8 T-cell suppressive capacity in two SIVcpz-infected chimpanzees. *J Med Primatol: 30:243-53.*

Openshaw, P.; Murphy, E.; Hosken, N.; Maino, V.; Davis, K.; Murphy, K.; - O'Garra, A. (1995). Heterogeneity of intracellular cytokine synthesis at the single-cell level in polarized T helper 1 and T helper 2 populations. *J. Exp. Med: 182:1357–67.*

Page, R.D. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*; *12: 357-358.*

Paradela, A.; Alvarez, I.; Garcia-Peydro, M.; Sesma, L.; Ramos, M.; Vazquez, J. and Lopez De Castro, J. (2000). Limited diversity of peptides related to an alloreactive T cell epitope in the HLA-B27-bound peptide repertoire results from restrictions at multiple steps along the processing-loading pathway. *J Immunol; 164:329-37.*

Parker, S.; Sadlon, T. and Gordon, D. (1995). Enhancement of NK cell-mediated antibody-dependent lysis of recombinant gp120-coated CD4 cells by complement. *J. Infect. Dis : 171:186–89.*

Parker, K.; Bednarek, M. and. Coligan, J. (1994). Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol: 152:163.*

Paliard, X. ; de Waal Malefyt, R. ; Yssel, H.; Blanchard, D.; Chretien, I. ; Abrams, J.; de Vries, J.; Spits H. (1988). Simultaneous production of IL-2, IL-4, and IFN-γ by activated human CD4$^+$ and CD8$^+$ T-cell clones. *J. Immunol: 141:849–55.*

Paul, W. (1998). Editor of Fundamental Immunology,Lippincott, Williams & Wilkins, Philadelphia, PA.

Preston, (1988). Fidelity of HIV-1 reverse transcriptase. *Science: 242:1168-71.*

Price, D.; Goulder, P. and Klenerman, P, (1997). Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc Natl Acad Sci USA: 94:1890–5.*

Putkonen, P.; Thorstensson, R. and Ghavamzadeh L, (1991). Prevention of HIV-2 and SIVsm infection by passive immunization in cynomolgus monkeys. *Nature: 352:436–8.*

Ramsay, J.; Richardson, J.; Carter, Y.; Davidson, L. and Feder, G.(2002). Should health professionals screen women for domestic violence? *Systematic review: 325:314.*

Rammensee, H.; Bachmann, J.; Emmerich, N.; Bachor, O.and Stevanovic, S.(1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics; 50:213-19.*

Robbins, J.; Schneerson, R. and Szu, S. (1995). Perspective: hypothesis: serum IgG antibody is sufficient to confer protection against infectious diseases by inactivating the inoculum. *J. Infect. Dis: 17: 1387–98.*

Robbins, J.; Schneerson, R. and Szu, S. (1999). Perspective: hypothesis: serum IgG antibody is sufficient to confer protection against infectious diseases by inactivating the inoculum. *J. Infect. Dis: 171: 1387–98.*

Robertson, D; Anderson, J; Carr, A.; Foley, B.; Funkhouser, R.; Gao, F.; Hahn, B.; Kalish, M.; Kuiken, C.; Learn, G.; Leitner, T.; McCutchan, F.; Osmanov, S., Peeters, M.; Pieniazek, D.; Salminen, M.; Sharp, P.; Wolinsky, S. and Korber, B. (2000). HIV-1 nomenclature proposal. *Science: 288:55-6.*

Rook, G.; Stanford, J.; Resmini, C.; Abbas, A. and Wolf, S. (1995). Adjuvants, endocrines and conserved epitopes; factors to consider when designing therapeutic vaccines. Int J Immunopharmacol;17:91-102.

Rowland-Jones, S.; Pinheiro, S.; Kaul, R.; Hansasuta, P.; Gillespie, G.; Dong, T.; Plummer, F.; Bwayo, J.; Fidler, S.; Weber, J.; McMichael, A. and Appay, V. (2001). How important is the 'quality' of the cytotoxic T lymphocyte (CTL) response in protection against HIV infection? *Immunol Lett: 79:15-20.*

Rowland-Jones, S.; Dong, T.; Fowke, K.; Kimani, J.; Krausa, P.; Newell, H.; Blanchard, T.; Ariyoshi, K.; Oyugi, J.; Ngugi, E.; Bwayo, J.; MacDonald, K.; McMichael, A. and Plummer, F. (1998). Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in Nairobi. *J. Clin. Invest: 102:1758-65.*

Rowland-Jones, S.; Dong, T.; Dorrell, L.; Ogg, G.; Hansasuta, P.; Krausa, P.; Kimani, J.; Sabally, S.; Ariyoshi, K.; Oyugi, J.; MacDonald, K.; Bwayo, J.; Whittle, H.; Plummer, F. and McMichael, A. (1999). Broadly cross-reactive HIV-specific cytotoxic T-lymphocytes in highly-exposed persistently seronegative donors. *Immunol Lett: 66:9-14.*

Rowland-Jones, S.; Sutton, J.; Ariyoshi, K. (1995). HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women. *Nat Med: 1:59–64.*

Rowland-Jones, S. and McMichael, A. (1995). Immune responses in HIV-exposed seronegative: have they repelled the virus? *Curr Opin Immunol: 7:448-55.*

Rosenberg, E.; Billingsley, J.; Caliendo, A.; Boswell, S.; Sax, P.; Kalams, S. and Walker, B. (1997). Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia. *Science: 278:1447-50.*

Rotzschke, O.; Falk, K.; Neefjes, J.; Hoglund, P.; Heemels, M.; Bastin, J.; Schumacher, T.; Townsend, A. and Karre, K. (1991). Exact prediction of a natural T cell epitope. *Eur J Immunol; 21: 2891-4.*

Sagar, M.; Lavreys, L.; Jared, M.; Richardson, B.; Barbra, A.; Mandaliya, K.; Ndinya-Ahola, J. and Kreiss, J. (2004). Identification of modifiable factors that affect the genetic diversity of the transmitted HIV-1 population. *AIDS: 18: 615-619.*

Saha, K (2001). Isolation of primary HIV-1 that target CD8+T lymphocytes using CD8 as a receptor. *Nat Med: 7: 65-72.*

Saitou, N. and Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol; 4: 406-25.*

Salminem, M; Koch, C; Sanders-Buell, E; Burke D and Carr J. (1995). Recovery of virtually full- length HIV-1provirus of diverse subtypes from primary virus cultures using PCR. *Virol: 213:80-6.*

Scala, G.; Chen, X.; Liu, W.; Telles, J.; Cohen, O.; Vaccarezza, M.; Igarashi, T. and Fauci, A. (1999). Selection of HIV-specific immunogenic epitopes by screening random peptide libraries with HIV-1-positive sera. *J Immunol: 162:6155-61.*

Scott, P. (1993). Selective differentiation of $CD4^+$ T helper cell subsets. *Curr. Opin. Immuno:* 5:391–97.

Sambrook, J.; Frisch, E.; Maniatis, T. (1989). Molecular cloning: A laboratory Manual, Second edition. Cold spring Harbor laboratory, New York.

Seth, A.; Ourmanov, I.; Evans, D. and O'Connor, D. (1998). Recombinant modified vaccinia virus Ankara-simian immunodeficiency virus gag pol elicits cytotoxic T lymphocytes in rhesus monkeys detected by a major histocompatibility complex class I/peptide tetramer. *Proc Natl Acad Sci U S A; 95: 10112-6.*

Sethi, K.; Omata, Y. and Schneweis, K. (1983). Protection of mice from fatal herpes simplex virus type 1 infection by adoptive transfer of cloned virus-specific and H-2-restricted cytotoxic T lymphocytes. *J Gen Virol: 64:443–7.*

Sette, A. and Sidney, J. (1999). Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics: 50, 201–212.*

Sharp, P.; Sugden, B.; and Sambrook, J. (1973). Detection of teo restriction Endonuclease activities in Haemophilus parainfluenza using analytical agarose. *J.Biochem: 12:3055-57.*

Shiver, J.; Fu T.; Chen, L.; Casimiro, D.; Davies, M.; Evans, R.; Zhang, Z.; Simon A.; Trigona, W.; Dubey, S.; Huang, L,; Harris, V.; Long, R.; Liang, X.; Handt, L.; Schleif, W.; Zhu, L.; Freed, D.; Persaud, N.; Guan, L.; Punt, K.; Tang, A.; Chen, M.; Wilson, K.; Collins, K.; Heidecker, G.; Fernandez, V.; Perry, H.; Joyce, J.; Grimm, K.; Cook, J.; Keller, P.; Kresock, D.; Mach, H.; Troutman, R.; Isopi, L.; Williams, D.; Xu, Z.; Bohannon, K.; Volkin, D.; Montefiori, D.; Miura, A.; Krivulka, G.; Lifton, M.; Kuroda, M.; Schmitz, J.; Letvin, N.; Caulfield, M.; Bett , A.; Youil, R.; Kaslow, D. and Emini EA. (2002). Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. *Nature:17; 415:331-5.*

Siegal, F. and Spear, G. (2001). Innate immunity and HIV. *AIDS: 15:S127-37.*

Siliciano, R.; Lawton, T. and Knall C. (1988). Analysis of host-virus interactions in AIDS with anti-gp120 T cell clones: effect of HIV sequence variation and a mechanism for CD4+ cell depletion. *Cell: 54:561–75.*

Slobod, K..; Lockey, T.; Howlett, N.; Srivinas, R.; Rencher, S.; Freiden, P.; Doherty, P. and Hurwitz, J. (2003). Clinical Evaluation of the First Multi-Envelope HIV-1 Vaccine: a 23 Envelope Recombinant Vaccinia Virus Cocktail. AIDS Vaccine 2003 Meeting New York, NY.

Spira, S.; Wainberg, M.; Loemba, H.; Turner, D. and Brenner, B. (2003). Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *J.antimicrib. Chemother: 51:229-240.*

Sölder, B.; Reisinger, E.; Koefler, D.; Bitterlich, G.; Wachter, H. and Dierich, M.- 1989. Complement receptors: another port of entry for HIV. *Lancet: 2:271–72.*

Soriano, V.; Gomes, P.; Heneine, W.; Holguin, A.; Doruana, M. and Antunes, R. (2000). Human immunodeficiency virus type 2 (HIV-2) in Portugal: clinical spectrum, circulating subtypes, virus isolation, and plasma viral load. *J Med Virol: 61: 111-6.*

Stout R, Bottomly K. (1989). Antigen-specific activation of effector macrophages by IFNγ producing (Th1) T cell clones. Failure of IL4 producing (Th2) T cell clones to activate effector function in macrophages. *J. Immunol: 142:760–65.*

Swain, M. and Huston, G. (1990). IL4 directs the development of Th2 like helper effectors. *J. Immunol: 145:3796–806.*

Sullivan, N.; Sun, Y.; Sattentau, Q.; Thali, M.; Wu, D.; Denisova, G.; Gershoni, J.; Robinson, J.; Moore, J. and Sodroski J. (1998). CD4-Induced conformational changes in the human immunodeficiency virus type 1 gp120 glycoprotein: consequences for virus entry and neutralization. *J Virol: 72:4694-703.*

Taube, R.; Fujinaga, K.; Wimmer, J.; Barboric, M. and Peterlin, B. (1999). Tat transactivation: a model for the regulation of eukaryotic transcriptional elongation. *Virology: 264, 245–53.*

Tatt, I.; Barlow, K.; Nicoll, A.; Clewley, J.; (2001). The public health significance of HIV-1 subtypes. *AIDS: 15:S59-71*

Takahashi, K.; Dai, L.; Fuerst, T.; Biddison, W.; Earl, P.; Moss, B. and Ennis, F.; (1990). Specificlysis of human immunodeficiency virus type 1-infected cells by a HLA-A3.1-restricted CD8+ cytotoxic T-lymphocyte clone that recognizes a conserved peptide sequence within the gp41 subunit of the envelope protein. *Proc. Natl. Acad.Sci: 88:1027-29.*

Takebe, Y.; Kusagawa, S. and Motomura, K. (2004). Molecular epidemiology of HIV: tracking AIDS pandemic. *Pediatr Int: 46:236-44.*

Tartaglia, J.;Cox, W. I.;Taylor, J.;Perkus, M.;Riviere, M.;Meignier, B.;Paoletti, E.; Tartaglia, J., W. I. Cox, (1992). Highly attenuated poxvirus vectors. *AIDS Res Hum Retrovir; 8: 1445-7.*

Termin, H.M. (1991). Sex recombination in retroviruses. *Trends Genet: 7:71-74.*

Thali, M.; Moore, J.; Furman, C.; Charles. M.; Ho, D.; Robinson, J. and Sodroski, J. (1993). Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J Virol: 67: 3978-88.*

Thomson, M.; Perez-Alvares, L. and Najera, R. (2002). Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect Dis: 2:461-471.*

Thompson, J.; Higgins, D. and Gibson, T (1994). ClustalW improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gag penalties and weight matrix choice. *Nucleic. Acids. Res; 22: 4673-4680.*

Tirado, G.; Jove, G.; Kumar, R.; Noel, R.; Reyes, E.; Sepulveda, G.; Yamamura, Y. and Kumar (2004). Compartmentalization of drug resistance-associated mutations in a treatment-naive HIV-infected female. *AIDS Res Hum Retroviruses: 20:684-6.*

Triques, K.; Bourgeois, A.; Vidal, N.; Mpoudi-Ngole, E.; Mulanga-Kabeya, C.; Nzilambi, N.; Torimiro, N.; Saman, E.; Delaporte, E. and Peeters, M. (2000). Near-full-length genome sequencing of divergent African HIV type 1 subtype F viruses leads to the identification of a new HIV type 1 subtype designated K. *AIDS Res Hum Retroviruses: 20:16:139-51.*

Tsomides, T.; Aldovini, A.; Johnson, P.; Walker, B.; Young, R. and Eisen, H. (1994). Naturally processed viral peptides recognized by cytotoxic lymphocytes on cells chronically infected by human immunodeficiency virus type 1. *J. Exp. Med: 180:1283M-P2.*

UNAIDS/WHO AIDS epidemic update December 2003. Website: www.unaids.org.

Vaishnav, Y. and Wong-Staal, F. (1991). The biochemistry of AIDS. *Annu. Rev. Biochem: 60:577-630.*

Vidal, N.; Peeters, M.; Mulunga-Kabeya, C.; Nzilambi, N.; Robertson, D. and Ilunga, W. (2000). Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J. Virol; 74: 10498-507.*

Villacres, M.; Zuo, J.; Bergmann, C. (2000). Maintenance of CD8(+) T-cell memory following infection with recombinant sindbis and vaccinia viruses. *Virol; 270: 54-64.*

Wang, L.; Chen, J.; Gelman, B.; Konig, R. and Cloyd, M. (1999). A novel mechanism of CD4 lymphocyte depletion involves effects of HIV on resting lymphocytes: induction of lymph node homing and apoptosis upon secondary signaling through homing receptors. *J. Immuno: 162: 268–276.*

Walker, E.; Preston, R.; Post, J.; Ehrlich, G.; Kalbfleisch, J. and Klingman, K. (1998). Genetic diversity among strains of Moraxella catarrhalis: analysis using multiple DNA probes and a single-locus PCR-restriction fragment length polymorphism method. *J. Clin. Microbiol: 36:1977-83.*

Walker, C. (1993). Non-cytolytic control of HIV replication by CD8+ T cells. *Semin Immunol: 5:195-201.*

Walker, C.; Thomson-Honnebier, G.; Hsueh, F.; Erickson, A.; and Levy, J. (1991). CD8+ T cells from HIV-1-infected individuals inhibit acute infection by human and primate immunodeficiency viruses. *Cell Immunol: 15: 137:420-8.*

Wee, E.; Patel, S.; McMichael, A. and Hanke, T. (2002). A DNA/MVA-based candidate human immunodeficiency virus vaccine for Kenya induces multi-specific T cell responses in rhesus macaques. *J. Gen Virol; 83: 75-80.*

Wong, S.Y.; Woo, C.; Luk, W. and Yuen, K. (1997). Nucleic acid technology and infectious diseases. *Hong Kong Med J: 3:179-185.*

Williams, N.; Engelhard, V.; (1997). Perforin-dependent cytotoxic activity and lymphokine secretion by CD4+ T cells are regulated by CD8+ T cells. *J Immunol: 159:2091-9.*

Williams, M.; Peh, C. A.;Purcell, A. W.; McCluskey, J.;Elliott, T. (2003). Bioinformatics tools for identifying class 1 restricted epitopes. *Immunit; 16: 509-20.*

Wilson, C.; Kalams, S.; Wilkes, B.; Ruhl, D.; Gao, F.; Hahn, B.; Hanson, I.; Luzuriaga, K.; Wolinsky, S.; Koup, R.; Buchbinder, S.; Johnson, R. and Walker, B. (1997). Overlapping epitopes in human immunodeficiency virus type 1 gp120 presented by HLA A, B, and C molecules: effects of viral variation on cytotoxic T-lymphocyte recognition. *J. Virol; 71:1256–1264.*

Wilson, C. C., B. Palmer, Koup, R.; Buchbinder, S.; Johnson, R. and Walker, B. et al. (2001). Identification and antigenicity of broadly cross-reactive and conserved human immunodeficiency virus type 1-derived helper T-lymphocyte epitopes. *J Virol 75(9): 4195-207.*

Wilson, S.; Pedersen, S.; Schmitt, D.; Kieny, M.; McNeil, J. and D. S. Burke. (1998). Cross-clade envelope glycoprotein 160-specific CD8+ cytotoxic T lymphocyte responses in early HIV type 1 clade B infection. *AIDS Res Hum Retroviruses; 14: 925-37.*

Wyand, M.; Manson, D.; Montefiori, J.; Lifson, R.; Johnson, and Desrosiers, R. (1999). Protection by live, attenuated simian immunodeficiency virus against heterologous challenge. *J. Virol; 73:8356–63.*

Wyatt, R. and Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens and immunogens. *Science: 280:1844-1888.*

Xiang, S.; Doka, N.; Choudhary, R.; Sodroski, J. and Robinson, J. (2002). Characterization of CD4-induced epitopes on the HIV type 1 gp120 envelope glycoprotein recognized by neutralizing human monoclonal antibodies. *AIDS Res.Hum. Retroviruses: 18: 1207–17.*

Zevering, Y. and Khamboonruang, C. (1994). Natural amino acid polymorphisms of the circumsporozoite protein of Plasmodium falciparum abrogate specific human CD4+ T cell responsiveness. *Eur J Immunol; 24: 1418-25.*

Zhang, C.; Fu, S.; Xue, Y.; Wang, Q.; Huang, X.; Liu, A.; Ma, L.; Shi, Z.; Zhang, Y.; Cheng, W.; Ai, Q.; Huang, C.; Chen, B.; Kang, X.; Sun, Y.; Zhang, G. and Li, P. (2002). Distribution of the CCR5 gene 32 –base pair deletion in 11 Chinese populations. *Anthropol. An: 60: 261-71.*

Zhu, T.; Wang, N.; Carr, A.; Nam, D.; Moor-jankowski, and Cooper, D. (1996). Genetic characterization of human immunodeficiency type 1in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. *J. Virol: 70:3098-3107.*

Zinkernagel, R. and Hengartner, H. (2004). On immunity against infections and vaccines: credo Review.*Scand J Immunol; 60: 9-13.*

Zinkernagel, R. and Doherty, P. (1997). The discovery of MHC restriction. *Immunol Today; 18: 14-7.*

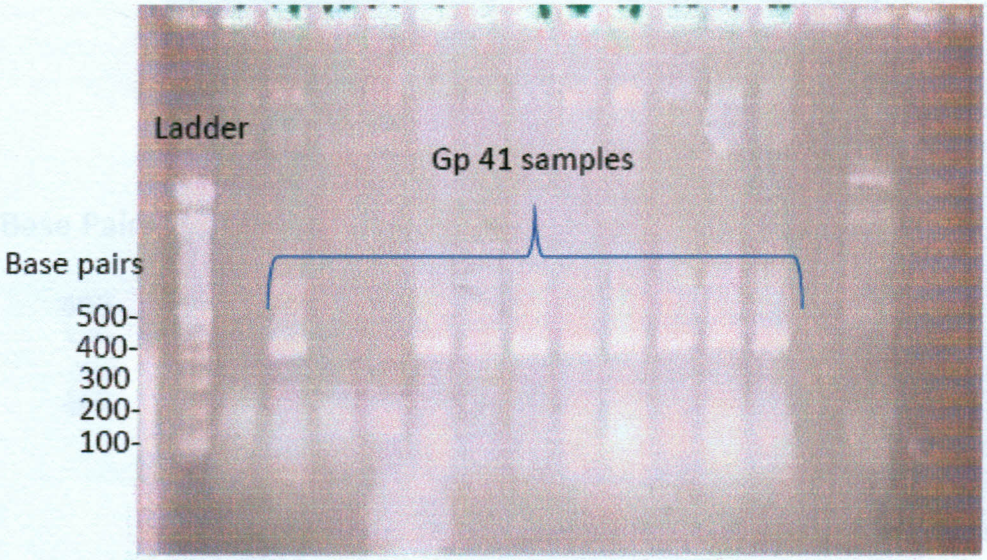
Zolla-Pazner, S. (2004). Identifying Epitopes of HIV-1 that induce protective antibodies. *Nat. Rev. Immunol: 4:199-210.*

Zolla-Pazner, S.; Xu, S.; Burda, A.; Duliege, J.; Excler, M. and Clements-Mann, M. (1998). Neutralization of syncytium-inducing primary isolates by sera from human immunodeficiency virus (HIV)-uninfected recipients of candidate HIV vaccines. *J. Infect. Dis: 178:1502–06.*

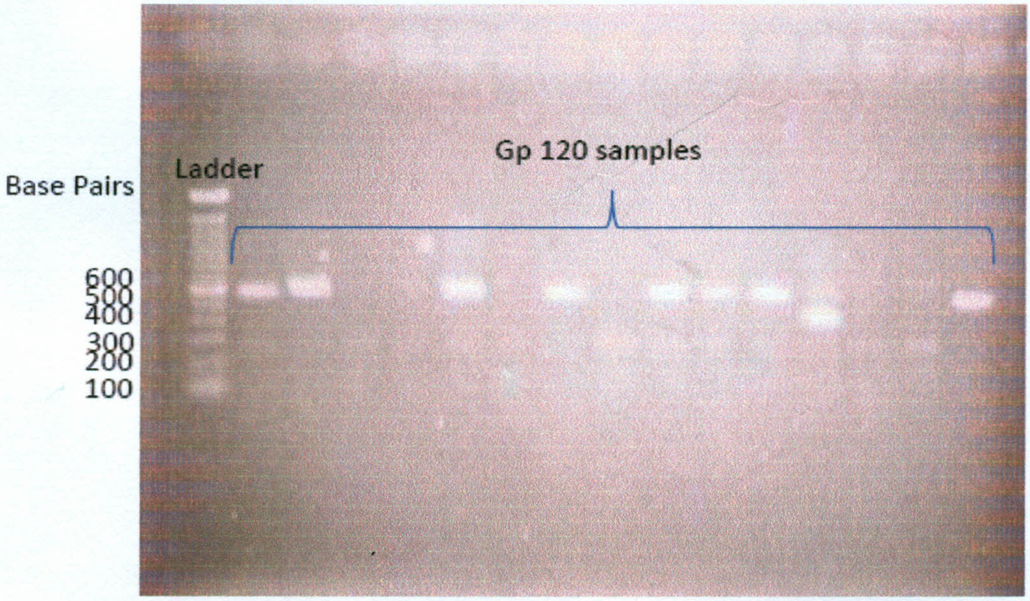
Zwick, M.; Wang, M.; Poignard, P.; Stiegler, G.; Katinger, H.; Burton, D. and Parren, P.; (2001). Neutralization synergy of human immunodeficiency virus type 1 primary isolates by cocktails of broadly neutralizing antibodies. *J Virol: 75: 12198-208.*

APPENDIX I

1. **Representative Gel electrophoresis of samples amplified by PCR using *gag, gp 41* and *gp 120* primers indicating the sizes of the product**

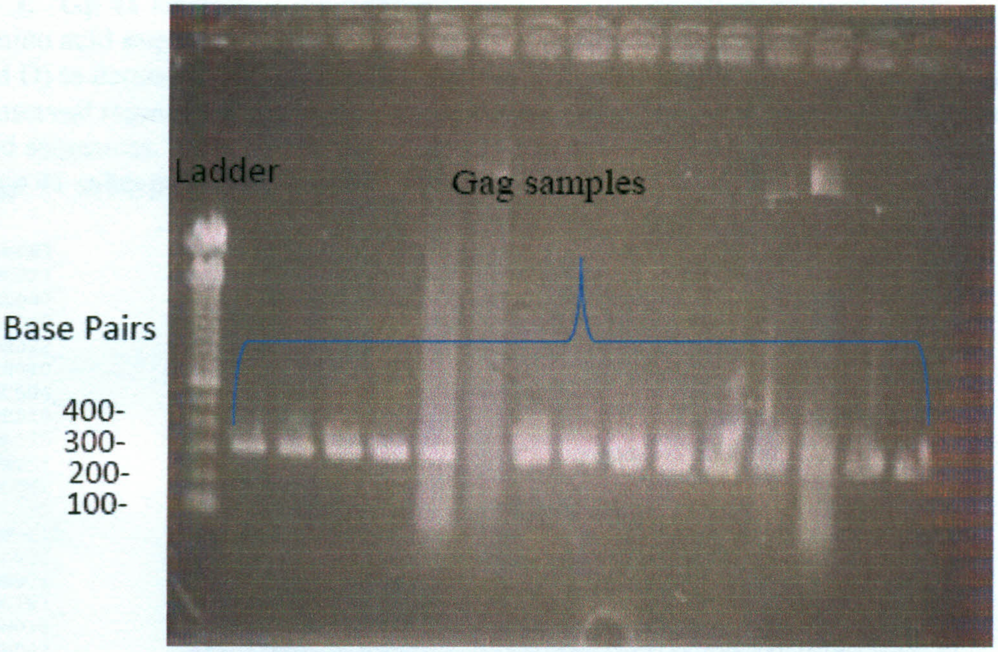


a) Gp 41 gel products were approximately 400 base pairs in length as shown



b) ***Gp 120* gel products**

The products from the gag region were approximately 600 base pairs in length



c) Gag gel products were approximately 300 base pairs in length

APPENDIX II

1. Gp 41 clustalw alignment

Amino acid sequence from gp 41 subtypes were aligned according to subtypes (A1, C and D) to determine if they were conserved. The areas marked in red colour indicate conserved regions. While blank area indicate regions with high variability in the amino acid sequences.

A)gp 41 subtype A1 alignment

```
MYDH063_        ------FEAIEA-QQHLLKLT-----VWGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 47
TLHC024_        -GGLTFAEGYRG-STASVETH-----GLGHSTAPGKILALER-YLKDQQLLGIWGCSGKL 52
MYSL030_        ----------KA-QQHMLQLT-----GLGHKQLQTRVLALER-YLKDQQLLGIWGCSGKL 43
MADH009_        -GAYNLWK-AQRLNNICCKLT-----VWGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 52
MYDH022_        -GAYNLWK-AQRLNNICCKLT-----VWGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 52
MYDH040_        ------LR-LSRLNNICCKLT-----VWGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 47
TLHC004_        ----NLFGGLSRLNNICSNSR-----SRGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 50
MYDH019_        -----LLKVLEA-QQQLLRLT-----VWGIKQLQARVLALES-YLRDQQLLGIWGCSGRL 48
MYDH015_        ---NFVGRLLEA-QQHMLRSR-----SRGIKQLQARVLALER-YLRDQQLLGIWGCSGKL 50
TKMH014_        -------EGIEA-QQHLLKLT-----VWGIKQLQARVLAVER-YLRDQQLLGIWGCSGKL 46
TLHC002_        LGDQNLWRAIEA-QQQMLRLT-----VWGIKQLQARVLAVER-YLKDQQLLGIWGCSGRL 53
MYSL033_        -WGHNLLRAIEG-STTSVETH-----SLGHKQLQARVLAVER-YLKDQQLLGIWGCSGKL 52
MYDH0020_       LEHYNFAGGLQRLNNICYKLT-----VWGIKQLQARVLAVER-YLKDQQLLGIWGCSGKL 54
MYSL032_        ----ICWRAIEA-QQHLLKLT-----VQGHQTAPGKSPRLWKDTLRDQQLLGIWGCSGKL 50
MYDH034_        ---------IEG-STTSVETY-----CLGHYNSSRQESWLWKDTSRINNSSGIWGCSGKL 45
TLHC101_        VFGGIEFVRLRG-STTSVETH-----GLGHSNSSRHCVLAVERYLKDQQLLGIWGCSGKL 54
MADH015_        --GDIIFGGLSRLNNICCNSR------SGALNSSRHSVLAIERYLRDQQLLGIWGCSGKL 52
MYDH042_        -------CGYRRLNNICRSSR------SGALNSSRQESWLSKDTSEDQQLLGIWGCSGKL 47
MYDH025_        ------LCGLYMLNNISLQTHGSRGHYTAPCKIPRAGGKDTWMGSTAPRNLGAGSGKTHL 54
TLHC037_        ---------LYRLNNICC-TH-SLGHYTAPGKSPGCG--KIPKGSTAPRNLGL-LWKTHL 46
MYDH026X_       -LLGPAIFGGCSSTTSVETHG------LGALNSSGQESLLWK-DTYGINSSSEFGAALES 52
MADH010_        -----NSVGLRGSTAYVETHS------LGHKQLQAR-SWLWK-DTERINSSYEFGAALED 47
MYDH038_        -------SRLSRLNNICCNSQ------SGALNSSRQESWLWK-DTYRINSSSEFGAALEN 46
MYDH026_        -----NLEGIEAQQHLLKLTV------WGIKQLRARVLAVER-YLRDQQLLGIWGCSG-N 47
MYDH051_        LFGARILXGLELTTSVAIHSL-------GHKQLQARVLGSWKDTSRDQQLLGIWGCSGKL 53
MYDH028_        ----------VEAQQHLLKLT-----VWGIKQLQARVLAVER-HLKDQQLLGIWGCSGKL 44
TLHC106_        --GHTICCRYRGSTAMVGDSR-----SGGIKQLQARVLALER-TYGINSSSVFGGCSGKL 52
MADH006_        ------LTIYAQHHLVKLTCR------ALNISMAMMPMLWEMIPKWSTAPMEFRGTLGKL 48
TLHC007_        ------LTIYAQHHLVKLTCR------ALNISMAMMPMLWEMIPKWSTAPMEFRGTLGKL 48
MYDH024_        ----FWGQRFGGSRQHMLNHG------LGHKQLQTRVLGYRK-IPKGSTAPRDLGLLWKT 49
TLHC112_        -----ILRALEAQQQLLNSRS------GGIKQLQARVLAVGKIPKGSTAPRNLGAALENS 49
                :
```

```
MYDH063_        ICTTNVPWNSSWS-NKTQGEIWD-NMTWLQWD---KRNLAITQISIYTLLEESQNQQEKN 102
TLHC024_        ICTTNVPWNSSWS-NKTQNEIWE-NMTWLQWD---KEIENYTQI-IYTLLEESQNQQEKN 106
MYSL030_        ICTTAVPWNMSWS-NKSQEDIWD-NMTWMQWD---REISNYTGT-IYRLLEESQNQQERM 97
MADH009_        ICTTAVPWNSSWS-NKTQDEIWG-NMTWLEWD---KEIDNYTQI-IYNLLEESQNQQERN 106
MYDH022_        ICTTAVPWNSSWS-NKTQDEIWG-NMTWLEWD---KEIDNYTQI-IYNLLEESQNQQERN 106
MYDH040_        ICTTAVPWNTSWS-NKSQSEIWD-NMTWMQWD---KEISNYTQL-IYELIEESQNQQERN 101
TLHC004_        ICTTNVPWNSSWS-NKSYNEIWD-NMTWLQWD---KEVSNYTDI-IYKLIEESQNQQEKN 104
MYDH019_        ICTTNVPWNSSWS-NKSQDEIWG-TLTWQQWE---KEISNYTEI-IYSLIEESQNQQEKN 102
MYDH015_        ICTTNVPWNSSWS-NKSLEKIWG-NMTWQQWD---KEIDNYTTL-IYSLLEESQNQQEKN 104
TKMH014_        ICTTNVPWNSSWS-NKSFTEIWD-NMTWLQWD---KEINNYTET-IYRLIEESQNQQEKN 100
TLHC002_        ICPTNVRWNSSWS-NKSYDEIWD-NMTWLQWD---KEISNYTQI-IYNLLEESQNQQEKN 107
MYSL033_        ICPTNVPWNSSWS-NKSQSEIWD-NMTWLQWD---REINNYTNL-IYSLIEESQNQQEKN 106
MYDH0020_       ICPTNVPWNSSWS-NKSLDAIWD-NMTWLQWD---REIDQYTQI-IYSLIEESQYQQEKN 108
MYSL032_        ICPTNVPWNHSWS-NKSYSEIWD-NMTWLQWD---REISNYTNI-IYQLIEESQNQQEKN 104
MYDH034_        ICTTNVPWNSSWSHNKTMDEIWE-NMTWLQWD---KEISNYTQL-IYSLIEESQIQQEKN 100
TLHC101_        ICPTNVPWNSSWS-NKSQEEIWG-NMTWLQWD---KEIDNYTQL-IYSLIEESQIQQEKN 108
MADH015_        ICTTDVPWNSSWS-NKSMHEIWD-NMTWLEWD---KEIGNYTDI-IYKLIEDSQNQQEKN 106
MYDH042_        ICTTNVPWNSSWS-NKSLEDIWD-NMTWLQWD---REISNYTQL-IYTLIEESQNQQEKN 101
MYDH025_        HHSW-CPGNSSWEVIKLRMEIWG-NMD-LGYNG-IKKLGNYTQV-IYNLIEESQNQQEKN 109
TLHC037_        HHS--CALDTSWS-NKSEADIWN-NLTWLQWD---KEISNYTDI-IYKLIEESQNQQEKN 98
MYDH026X_       SSAPLMCPGTLVGVINLKRRYXE-NMTWLQWD---REISNYTDI-IQELLEVSQNQQEKN 107
MADH010_        SSAPLMCPGTPVGVINLIMKYWD-NMTWLEWD---KEIGNYSDI-IYRLIEESQNQQEKN 102
MYDH038_        SSAPLMCLGTLVGVINLYMKYGR-TCPGCNGI---KKLTIIHKQ-YINSLKNRRTNRKRM 101
MYDH026_        SSAPLMCPGTLVGVINLKRRYGR-TDLVTMGY---GKLAITQTR-YMGLLEVSQNQQEKN 102
```

```
MYDH051_        ICTTNVPWNSSWS---LNLRVRY-GTTCPGCNG-IKKLAITHRQ-YINYLKNRKTSKEKN 107
MYDH028_        ICTTNVPWNSSWS-NKSQELRYG-ITDLAAMG--YRNYQL-HRHNISTNCRIAEPAGKEC 99
TLHC106_        ICTTNVPWNSSWE-YYNPGGPYG-TTCPAAMG--LKKLAITHRSYMGHFEGTQNQPGKEC 108
MADH006_        ICTTYWSPGNSKLGVIILKMDDM-GIPWPGSNGDEMKLAITQTYYIIYLEESQEPSGKMN 107
TLHC007_        ICTTYWSPGNSKLGVIILKMDDM-GIPWPGSNGDEMKLAITQTYYIIYLEESQEPSGKMN 107
MYDH024_        HLHHCCPWNSSWS--LNLKRIFG-YHDLDAVG---SRNSSLHRHNIQVACSLAKPAGRKC 103
TLHC112_        SAPLMALGTPNWSVKSLWRDFWESPWTCSNWDK--PKLAIYTGRPLYYISSEDFPLPSRE 107
                                                    .
```

```
MYDH063_        EQDLLALDKWAS--LVELVSHNTLLVVYKDIYNDVG-ALR--- 139
TLHC024_        EQDLLALDKWAS--LWNWFKHIKMAVVYKNIYNDSR-RLG--- 143
MYSL030_        KR---TYQHWTVGKICGIGQYIKLAVVYKDIHNDSREALRDP- 136
MADH009_        EQDLLALDQWAN---LWSWFDISHWLWIQDIHNEVR-RLG--- 142
MYDH022_        EQDLLALDQWAN---LWSWFDISHWLWIQDIHNEVR-RLG--- 142
MYDH040_        EQDLLAFDKWAS---LWNWVCI--YQIAVDIKIFIS-IVG--- 135
TLHC004_        EQDLLALASGQV---LWNWFSITNWLWISNIYMIVG-GLG--- 140
MYDH019_        EQELLALASGPN---LWNWLIYKMAVDIKIFIMIVG-GFEESK 141
MYDH015_        EQDLLALASGQI---CGIGCHIKMAVDIKIFIMIVG-GLG--- 140
TKMH014_        EQDLLALASGQI---CGIGLTSQNAMDIKIFIMIVG-ALRV-- 137
TLHC002_        EQDLLALDSGQV---CGVGLALTTGCGYKIFIMIVG-AWG--- 143
MYSL033_        EQDLLALDNGQI---CGIGLTYQIGCGTQDIYMIVG-GLG--- 142
MYDH0020_       EQDLLALDSGQA---CGIGLTYQIGCGTQRYLYDSR-RLRG-- 145
MYSL032_        EKDLLELDKWAN---LWNWFDISRWLWTQKYLLMIVGGLG--- 141
MYDH034_        EQDLLSLASGQG---LWN-WLRYQIAVDIKILYDSR-RL---- 134
TLHC101_        EQDLLALACGQI---YGIGFSISKWLVDIKIFIMIV-GA---- 143
MADH015_        EKELLELANG---QICGIGLTYHIGCGIYDIYMIVG-RLG--- 142
MYDH042_        EQELLALASG---QICGVGLTYQTGCGYKNIYMIVG------- 134
MYDH025_        EQDLLGIGQVG--KSVELVCHLKMAVVYKNIYMIVG-GLG--- 146
TLHC037_        EQDLLALDKW---QICGIGLKYQNGCGYKDIYMIVG------- 131
MYDH026X_       EQDLLALDKWQV---CGVGLTYQDAVGY-KIFIMIAGA----- 141
MADH010_        EKELLELDKWAS---LWNWFDISKWLWYIKIFIMIVGGLG--- 139
MYDH038_        SKTYCNWTNGQI---CGIGLTYQIGCGYKNIYNDVGAW----- 136
MYDH026_        EQDLLALDSGQV---CGVGRHNKMAVDIKIFIMMYEAWG---- 138
MYDH051_        EQDLLAFGQVGK--SVDLVCHIKIGLWYIKYLYWSEG------ 142
MYDH028_        TR-LIGIRQMGKFVAWFASQIAV---DIR-ILYDSRSLG---- 133
TLHC106_        TKHLLALDKVGKFCGNGLSIFKPGWGDIKKFIMMLEAL----- 146
MADH006_        GQGFIGIGTRWGKVWWELGWTSHRWLWISRIFIMVGRRLG--- 147
TLHC007_        GQGFIGIGTRWGKVWWELGWTSHRWLWISRIFIMVGRRLG--- 147
MYDH024_        KRFITHWTVWKN---LWNWFSITNWLWIYGIHNDRGALG---- 139
TLHC112_   IKGKT----------------------------------- 112
```

---

b) gp 41 Subtype C Alignment

```
MYDH0054_       ----LRVYEA-QQLIVETHSLGALSSSRQESSALERYLKDQQLLGIWGCSGNSSAPLLCL 55
MYDH059_        ----LEAIRG-ATAYVATHSLGHLSSSRQESWLSKDTYRINSSSGFGAALENSSAPLLCL 55
TLHC109_        PSIFAKGYRKRNSRMFATHGLGASNSSRRESWLWKDTSGINSSSGFGGCSGKLICTTNVP 60
MYDH014_        ----NLLRLLEAQQHMLKLTVWGIKQLQTRVLAIERYLQDQQLLGIWGCSGKLICTTAVP 56
MYDH007_        ----NLLRAIEAHSIQLQLTVWGIKQLQTRVLAIERYLQDQQLLGIWGCSGKLICTTAVP 56
MYDH004_        ----ICLRAIEAQQHLLQLTVWGIKQLQTRVLAIERYLKDQQFLGIWGCSGKLICTTSVP 56
MYDH008_        -------EGIEAQQHMLQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVP 53
MYDH018_        --------GIEAQQHVLQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVP 52
MYDH023_        ------GDYRRRNSIYSQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVP 54
MYDH055_        ------EGYRGATAIVATHGLGHPKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVP 54
MYDH039_        ------EGYICATSYVATHGLGH-YSLHARFLAIEIYLKDQQLLGIWGCSGKLICTTAVP 53
MYDH003_        ---NLFGGYRGASAYVATHGLGHLSSSRQESQAIERYLQDQQILGIWGCSGKLICTTNVP 57
MADH001_        --------PRGATAYSSNSHSRGIKQLQTRVLAIERYLKDQQLLGIWGCSGNSSAPLMCL 52
MYDH002_        --------PRGATAYSSNSHSRGIKQLQTRVLAIERYLKDQQLLGIWGCSGNSSAPLMCL 52
MYDH001_        -----ICCGIEAQQHMLQLTVYGIKQLQARVLALERYLRDQQLLGMYGCSRNSSAPLLCL 55
TKMH020_        ------WKAIEAQQHMLQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGNSSAPLMCL 54
MADH013_        -------------HIVATHSLGALSSSRQESLALERYLKDQQLLGIWGCSGKVVCTTAVP 47
MYDH005_        -------------HIVATHSLGALSSSRQESLALERYLKDQQLLGIWGCSGKVVCTTAVP 47
MYDH041_        ------LKVSKANNICCNSRFRGIKQLQTRVLALERYLKDQQLLGIWGCSGKLICTTAVP 54
TLHC014_        ---------FQRRNSVVATHSLGHKQLQTRVLAIERYLKDQQLLGIWGCSGKLICTTAVP 51
TKMH017_        --------TIEAHSICCSSRSGALSSSRRESWLLEKYLKDQQLLGIWGCSGKLICTTTVP 52
TKMH017X_       -------ETIERNSICCSSRSGALSNSRRESWAIEKYLKDQQLLGIWGCSGKLICTTTVP 53
MYDH006_        --SDFAKGLKGATALFASHGLRHLSSSRQAVLAIKKYLKNQPLLGISGCSCKTHLHHCCT 58
```

```
MYDH012_        -------GYRGATAMLQLTVW-GIKQLQDKSPSYRKIPKGSTAPRTLGLLWKTHLHHCCA 52
MYDH011_        -------EGYRGATAYSCNSRS-GALSSSRQESSLWKDTQRINSSSEFGAALENSSAPLLC 53
MYDH013_        ------EGYRGATAYSCNSRSRGIKQLQTRVLALGKIPKGSTAPRNLGLLWKTHLHHCCA 54
MYDH058_        -----FEAIEGATAYVA-THGLGHSAAPDKSPGYRKIPKGSTAPRDLGLLWKTHLHHCCA 54
MYDH027_        TGATICWTLRGATALLA-THGLGIKQLHARVLALERYLKDQQLLRDLGLLWKTHLHHCCA 59
                                                              .   :

MYDH0054_       GTLVGVIN-LS-MIFGRTCLGCSGIEKLSNYTNTIYKLLEDSQIQQDNNEKNFTCIGQVG 113
MYDH059_        GTLVGVIN-LK-MIFGRHDLDAVGYEKLSNYTDTIYKLLEDSQIQQENNEKELLAL-AVG 112
TLHC109_        WTLVGVIK-LK-MTFGIKRPGCSGIERLSNYTDIIYRLLEKSQNQQEQNEKDLLAL-ANG 117
MYDH014_        WNSSWSNK-TEKEIWETLTWMQWEREID-NYTEIIYGLLEVSQNQQEKNEQDLLALDKWQ 114
MYDH007_        WNSSWSNK-TDTEIWENMTWMQWDREIS-NYTEIIYELLEVSQNQQEKNEQDLLALANGK 114
MYDH004_        WNSSWSNR-SYNEIWDNMTWMQWDREIS-NYTNTIYRLLEDSQNQQEKNEQDLLALAVGK 114
MYDH008_        WNSSWSNK-SQGEIWENMTWMQWDREIN-NYTDTIYRLLEVSQNQQEENEKDLLALDSWK 111
MYDH018_        WNSSWSAK-NQTEIWNHTTWMEWDREIS-NYTDTIYRLLEDSQNQQDENEQ-LLALDKWK 109
MYDH023_        WNSSWSNK-SQDYIWDNMTWMQWDKEIS-DYTSIIYNLLEVSQNQQEENERELLALAVGI 112
MYDH055_        WNSSWSNK-SKEEIWNNMTWMQWNKRNS-YLHRHNIQVACRIANPAGNNEKDLLALDSWK 112
MYDH039_        WNSSWSNK-SKEEIWENMTWMQWDREIN-NYTDTIYRLLELSQNQQEENERDLLALDKWQ 111
MYDH003_        WNSTWSNK-SEKEIWDNMTWMQWDREIA-NYTTIIYELLEVSQNQQEQNEKELLALANGR 115
MADH001_        GILVGVIE-LM-MRFGRTRPGCNGRGKL-IITQASYMICLKSHKTSRRKMSKIYQHWTNG 109
MYDH002_        GILVGVIE-LM-MRFGRTRPGCNGRGKL-IITQASYMICLKSHKTSRRKMSKIYQHWTNG 109
MYDH001_        RTPVRVIN-LY-EGFKIIVPRCNRIKKL-IIIQTSYTIYLKSRKTSKKQMKKIYSLWQMA 112
TKMH020_        GILAGVTN-LR-RIFGITCPGCSGRKKL-AITQIQYMDCLKNRKTSRKSMNRSSSIGKWQ 111
MADH013_        WNISWSNK-SK-QDIWDKMTWMQWDKEL-AITQTQYMNCLRNRKSSRKTMKRIYQHWQLE 104
MYDH005_        WNISWSNK-SK-QDIWDKMTWMQWDKEL-AITQTQYMNCLRNRKSSRKTMKRIYQHWQLE 104
MYDH041_        WNNSWSNR-LK-MIFGKTRPGCSGIEKL-VITQAQYTGCLRSHKTSRKKMNKIYQHWQLE 111
TLHC014_        WNSSWSNK-SY-SDIWITCPGCSGIEKL-VIIQTQYTSCLKIT-IQQEKMKKIYQHSANG 107
TKMH017_        WNSSWSNR-SQDDIWNNMTWMQWEKEID-NYTDTIYKLLEVSAEPTGAKCKGTYCIGTNG 110
TKMH017X_       WNSSWSNR-SQGMKFGITCLGCSGRKKL-IIIQTQYTNCLKSHRTNRSKMKRTYWHWQMA 111
MYDH006_        LGLYWSLQ-SQKKIQNHITYMQWNKKIS-NYTGTIYKLLNES------------------ 98
MYDH012_        LERIAGVI-NHNMIFGITCLGCSGIEKLIITQTQYTSCLKNRKPSRNRMKK-IYSHWTVG 110
MYDH011_        LGIVVGVI-NLKGIFGITCLGWNGIEKLIITQTQYTGCLKNRKTSRKKMKK-IYYHWQLE 111
MYDH013_        VESSWSNK-SQGAIW-VTRLGWSAIEKLIITQTHYTGCLKNRKTSRNKLKRDLLALDSLT 112
MYDH058_        LELYLEK--NLKRKFGITWPGCSGIEKLVITQALIYTLLEESQNQQESNEKDLLALDSWK 112
MYDH027_        LGTVVGGNRTKDYIWDNMTWDAGGIEKLAITQTQYTGWLEGIANPAGNKEKDLLALDSWN 119

MYDH0054_       TILVELGSHNQTGCGYYNYSNDSRRLCGFQ 143
MYDH059_        -QSVELVKHNKL-LWIYKISYDSKEAW--- 137
TLHC109_        -KICGVGLTYQIAVDIKNFIMIVG-AWG-- 143
MYDH014_        -NLWLGLAY--QLLWISKHSNDIGGLRIQ- 140
MYDH007_        -ICGIGLAS--QIAVDIKFHNDVGGLG--- 138
MYDH004_        -SMELVSIT--IGCGYKNIHNDSRRLG--- 138
MYDH008_        -NLWSWFDIT-NWLWTKDFIMIVGGLG--- 136
MYDH018_        -ICGIGLTYP-IGYGHKDFIMMSEA----- 132
MYDH023_        -ICGIGLAYQSGWSILKIFIMIVGALG--- 138
MYDH055_        -ICGIGLTSQ-TGCGYKNFIMIVGA----- 135
MYDH039_        -NLWTWLTYQ-TGCGYKDIHNDSREAWG-- 137
MYDH003_        -ICGLVDIT---IAVDIKYSYDSRRLG--- 138
MADH001_        -KICGVGLTYQTGCGYKIFIMMDG------ 132
MYDH002_        -KICGVGLTYQTGCGYKIFIMMDG------ 132
MYDH001_        -KFIELVRHNTVAVDIKIFIMIVGGLG--- 138
TKMH020_        -NLWSW-FDITHCCGYKNIHNDVGG----- 134
MADH013_        -QSVELVDISKWLW-IYIFIMIVGGLG--- 129
MYDH005_        -QSVELVDISKWLW-IYIFIMIVGGLG--- 129
MYDH041_        -KSVELVSHNKLLWYIRIFIMIVGGLG--- 137
TLHC014_        -KICGVGWHNNCCG-YKTIHNDSRRLG--- 132
TKMH017_        -KPVELVCISRVLVGYKNIHNDRRSFE--- 136
TKMH017X_       -KPVELVCISRWLWIYEYYNDSRRL----- 135
MYDH006_        ------------------------------
MYDH012_        -KTCGVGLTYQIAVDIKIFIMIVGG----- 134
MYDH011_        -NLWNW-FDIQIAVDIKIFIMMSEL----- 134
MYDH013_        -NLGNWVDISKWLWYIKISLMIVGGFEG-- 139
MYDH058_        -NLWNW-FDITTGCGYKNFIMMYGAW---- 136
MYDH027_        -NLWNW-LAKQNGCGYKDIHNDSKGGLG-- 145
```

c) gp 41 Subtype D Alignment

```
MYDH057_        ------FAEGITRPTAYVALTVRGIKQLQTRVLAIER-YLKDQQLLGIWGCSGKLICTTA 53
TLHC034_        ------FAEGYRRRNTSVATHCLGHKQLQARVLAVES-YLRDQQLLGVWGCSGKHICTTA 53
MYDH062_        ---------RYRGATAYVRTHSLGHYTAPGKSPGCRKGYLKDQQLLGIWGCSGKLYLQPL 51
MYDH035_        -----ICCGAIRRRTSSVATHSLGRYTAPGKNPGCGK-NLKDQQLLGIWGCSGSHICTTT 54
TLHC009_        -----TFHEGYQRRSSFVATHSLGLSNSSKQESWLWK-GPSRSTAPRNLGFSGSTLFRTT 54
TKMH015_        -----NICLRAIEAQQLCCNSQSGALNSSRQESWLWK-TTQGPTAPRNLGFSGRHICTTN 54
MYDH0021_       FGASEFFCGYQRLNNICWKLTVWGIKQLQGTESWLWK-DTSGINSSSEFGVALENIFAPL 59
TLHC040_        -----IFCDYRGAT-ALLQLTVWGIKQLQ-ARSWLWK-ATYRINGSSEFGVALENTFAPL 52
MYDH009_        -----NIAEGIEAQQHCCNSQSGALNSSR-RESWLWK-DTYGINSSSGFGVALEDTFAPL 53
```

```
MYDH057_        -VPWNSSWSNKSY-QDIWENMTWMQWDREINNYTNTIYRLLGRFAKPAGRK-------CK 104
TLHC034_        -VPWNASWSNKSI-EEIWNNMTWMEWEREIENYTGVIYSLIEESQTQQEKNEQELLQLAN 111
MYDH062_        NVPLELLAGVTSLKKEIWNNMTWMQWEREIGNYTDTIYGLLEESQTQQEMN--------E 103
MYDH035_        -VPWNSTWSNRSL-GEIWNNMTWMQWEKEIDNYTGIIYSLIEASKPSRKRMNKNFWHWTN 112
TLHC009_        NVPWNSTWGNKSL-DKIWN-MTWIQWEKEIDNFPGIISSLIEASKPSRKGMTRNFWNWPM 112
TKMH015_        -VPWTIVGVINHLLEEIWGNMTWMEWEKEISSYTGVIYNLIEESQTQQEKMKKNYWNWTN 113
MYDH0021_       TCPGTLVG-VIKLMVRFGITCPGWSGKEKLTITQVYYIPYLKNRKPSKKKCTRTIGIGPM 118
TLHC040_        LCPGTLAG-VIDLRCHLARTSHGWSGKGRLTITQNSYTASLKNRKSSKKRMKKIYCNWTH 111
MYDH009_        MCHGTLAG-VIKLSSRIWDNMTWMEWEREIDNYTGLIYNLLEKSQAQQEEMNKTYCNWHM 112
                                :       . . .:          :
```

```
MYDH057_        GFTLHWDSWDNLWSWLYITNWLWEIKGYS---- 133
TLHC034_        GQVCGIGSHNKMAVDIKIFIMLQELCGFQ---- 140
MYDH062_        QELLPLDKWPNLWNWFDISHWLWIKNFIMKWRL 136
MYDH035_        GQACGIGLHTKVLWIYEYSYCCRRLWG------ 139
TLHC009_        GKLVELVSHTHGCGYKILSSCRSLG-------- 137
TKMH015_        GQVCGMVCHNKLAVDIKIFIMMSELWG------ 140
MYDH0021_       GKFVELVSITNGCGYKNIHNDSRRLG------- 144
TLHC040_        GQVCGLVSHNKCCGYKNIHNDVGGFG------- 137
MYDH009_        GKFVELVNITNWLWISKYSNDSRRLG------- 138
```

2.Gag ClustalW alignment

Amino acid sequence from gag subtypes were aligned according to subtypes (A1, C and D) to determine if they were conserved. The areas marked in red colour indicate conserved regions. While blank area indicate regions with high variability in the amino acid sequences.

a) Gag Subtype D alignment

```
MADH007_        ---------GFSPCQHFGHKTRTKGNPLETMSIG-SIKLSEP-SKLHRRSKTGCQKPCWS 49
MYLD005_        ---------GFSPCQHFGHKTRTKGNPLETMYIG-SIKLSEP-SKLHRRSKTGCQKPCWS 49
TKMH019_        --------QHFG------HKTGTKG-TLSRLCRS-VLSNSKSRASFTGCKKLDDRNLVGS 44
L7_             ------GMSDFSPASILDIRQGPKE-PFRDYVDL-VLSNSKSRASFTGSKKLDDRNLVGS 52
TKMH011_        ------------APSILDIRQGPKE-PFRDYVDR-FYKTLRAEQATQGGKKLDDCNLVGP 46
MYSL032_        ------------ALQHFGHKTGAKG-PFRDYVGPGFYKTLRAEQASQDVKNWDDRNLVGS 47
MYDH008X_       -----GLVR-FSPCQHFGHKTRTKG-TLSRLRRS-VLSNSETEQASQDVKNWMTETLL-V 51
MADH001_        -----GLVR-FSPCQHFGHKTRTKG-TLSRLRRS-VLSNSETEQASQDVKNWMTETLL-V 51
MADH012_        -----------LALPAFWTYDKAKG-TLSRLCRS-VLYNSKSRASFTSMYKIGCQKPCWS 47
L8_             -----------LALPAFWTSDKAKG-TLSRLCRS-VLSNSKSRASFTSMSKIGCQKPCWS 47
MYDH025_        -----SGLVRFTALQHFGHKTRTKG-PFRDYVGS-VLSNSKSCASFTRCKNLDCQKPCWV 53
MYDH057_        -----------LGPPAFWTSNKAKG-PFRRLCRIGSIKLSEP-SKLHRMSKIGCLETLLI 47
MYDH022_        PTCHFQNSCDVSPCSTLDIRQGPRN-PLETMVGP-VLSNSKSRASFTRWKTWMTETLL-G 57
TLHC014_        ---LLGXSKFXGRPSILGHKTRTKXTLXEDYVRS-ALSNSKSCAMXTXMQXIRMTDTF-- 54
```

```
MADH007_        KMLNPDCKTILKALGPALHSKKCCQHVRE-WGDPVIKQGFWLRQCAKQQISILLS---CC 105
MYLD005_        KMLNPDCKTILKALGPALHYKKCCQHVRE-WGDPVIKQGFWLRQCAKQQISILLS---CC 105
TKMH019_        KMQIQIVKLSSKHWDQGLHSKKCCQHVKE-WEDPVIKQGFWLRQCAKQQMQMLLY---CC 100
TKMH011_        KCEPRLYEYLKSIWGPAATLEEMMTACQG-VGGPSHKARVLACGNEPSIKCKCCY---ND 102
MYSL032_        RMLNPDCKTILKALGPVQPHSKKCCQHVR-EWGAQPSSKSFGCGNEPSIRCKCCY---ND 103
MYDH008X_       QNANPDCKTILKALGPGATLGEMMTACQG-VGGPAYG-KGFGCGN-CAKQQVGLL---SC 105
```

```
MADH001_            QNANPDCKTILKALGPGATLGEMMTACQG-VGGPAYG-KGFGCGN-CAKQQVGLL---SC 105
MADH012_            KMLNPDCKTILKALGPAATLEEMMTACQG-VGGPSHKARVLAEAMSQATQANTAV---MM 103
L8_                 KMLNPDCKTILKALGPAATLEEMMTACQG-VGGPSHKARVLAEAMSQATQANTAV---MM 103
MYDH025_            QNANPDCKTILKALGPGATLGRHDDSMSGESGGPSHKSKEFLACGNCSQATNGKCXSCCC 113
MYDH057_            QNAEPRLSNYLKSTVFRGYIGGKCCQHVKEWGGPSHKARVFGCGNCAKSIPPA------- 100
MYDH022_            PKCEPRLSNYLKSIGTRGYIGGTCCQLVSXXXGPSYSSKXXRXEXXXXXXIGXAX----- 112
TLHC014_            --XGPKCXPRLXXLSS-------------------------------------------- 68
```

```
MADH007_            REAILEAKR----ESLSVSTVARRPLARNCKVLGKRLCK--- 140
MYLD005_            REAILEAKR----ESLSVSTVARRPLARNCKVLGKRLCK--- 140
TKMH019_            REAILRAQR----KPLSVSTVAKEGTSQ--KIAGS------- 129
KMH011_            AERQFSGPK----EDHSVFSCGKEGTPSQKLYGPL------- 133
MYSL032_            AERQFSGPK----EKHSVFNCGKGGAHSKELRAPRKRLCNWG 141
MYDH008X_           CRKAFSGAK----EKLSMFTGQR-GPLSRKCRLLRKGCGN-- 140
MADH001_            CRKAFSGAK----EKLSMFTGQR-GPLSRKCRLLRKGCGN-- 140
MADH012_            QSGNFKAQG----KPLSVSTVAR-RHLARYCRVLEKRLCKWG 140
L8_                 QSGNFKAQG----KPLSVSTVAR-RHLARYCRVLEKRLCKWG 140
MYDH025_            KSAXLRXXX----KPLXVSTXGKEGFLARXCXASXE------ 145
MYDH057_            ------------------------------------------
MYDH022_            XXCCKKGXX--------XXVPGKTXN---------------- 130
TLHC014_            ------------------------------------------
```

b) Gag subtype A1 alignment

```
MADH011_            ----SSPCQHFGHKA-----RAKCT-LQKIMSDR-FFKTL--RAKQATQEVKGCIT-ETL 46
MYLDL03_            ----SSPCQHFGHKA-----RAKCT-LQKIMSDR-FFKTL--RAKQATQEVKGCIT-ETL 46
TKMH012_            ---FSQTVRFIALLAFWISNKGPKN-HSEIMYDR-FFKAL--RAEQATQEVKHCMT-ETL 52
MYDH037_            ---------LYSPVSILDIKQGPKE-PFRDYVDR-FFKVL--RARASHTGGKRLDDMKHC 47
MADH005_            --SFSDLVRFIRPVSILDIKQGPKE-PFRDYVGSGSLKLS--ELSKLHRRSKVGWT-ETL 54
MYDH028_            ----------DSPVSILDIRQGPKE-PFRDYVRSVLYNSQ--SLSKPHRMSKIGST-ETL 46
MYLD07_             -------------VSILGHKTRAKR-PFRDYVDTVLSNSQ--SCASYTGGKRLGCQ-KPC 43
MYDH038_            -------------VSILDIKQGPKE-PFRDYVDRILSNSQ--SCASYTRGKRMDRL-ETL 43
MADH010_            ----------PSGAQHFGHKTRAIC-NPSETMSTGSLKLS--EPNKLHKRSKLGST-ETL 46
TLHC023_            ---------GSLKLSEL---SKLHR-RYK--VG--------------MTG--------NI 23
MYSL033_            -------TSQILALQHIGYKTRAKR-PFRDYVGQ-VLSSS--QSCASHTGCKKLDD-RNI 48
TKMH010x_           --------------SHFGCKTRAKR-PFREYVDT-VF----------------------- 21
MYDH063             --------SLYMPSSHFGYKARSKR-PFRDYVDR-FFKTL--RARASYT-GGKRLD--CQ 45
MYDH0035_           ---------FLRPASILGYKTGAKR-PFRGLCRIGFFKTL--RSLKQATQGGKKXG--XT 46
MADH014_            ---------RCIALLASWISNKGQK-NPSEIMSIGSLKLS--ELSKLHRRSK-VGCQKHY 47
MYDH110_            ---------RCIALLASWISNKGQK-NPSEIMSIGSLKLS--ELSKLHRRSK-VGCQKHY 47
MADH004_            ---------RFLALLAFWISNKGQK-YPSEIMSIGSLKLS--ELSKLHRKSK-VGCQKHY 47
MYDH026-1_          ---LFNNSSDVSPCSHFGYKTRAKR-PFRDYVGSGSLKFS--ELSKLHRMSK-IGCQKHY 53
MYDH026-2_          ---LFNNSSDVSPCSHFGYKTRAKR-PFRDYVGSGSLKFS--ELSKLHRMSK-IGCQKHY 53
MYDH049_            -----------AP-SHFGYKTRAKR-TLQSIMSIGSLKFS--EPSKLHRRSK-VGCQKHY 44
MYLDX005_           -----------FSPASILVSNKGQK-NLSETMSIGSLKLS--ELSKLHRMSK-IGCQKHY 45
TLHC019_            ----FIRPASILDIK------QGPK-NPSEIMYDRVLQSP--QSCTSFTGGKRLDCQTHY 47
MYDH006_            ----FVRTSKMLALLVFWISNKGQK-NPSEIMSDRVLSNS--QSCASYTGGKRLGCQKHY 53
TLHC018_            -----------IALLAFWISNKGQK-NPSEIIVDRFFKVL--RAEQATQVGKRFGCQTHY 46
MYDH046_            -------SYVVYALLAFWISNKGQK-NPSEIMSIWFFKVL--RAEQATQEVKGSDRQKHY 50
MADH008_            ------PSKIFSPASILGYKNRGQK-TLQRLWSDRVLLKSQ-SCASYTGGKGVGXDKXLL 52
MYDH015_            ----LLKPVRFSPCSFLD-SNRGQK-TLQRLWWDRVLSNSQ-SLSKLHREVKGSDDRKHY 53
MADH003_            ----FLITVRLSPCSHFGYKTRAKR-TLQRLCR--------------------------- 28
TLHC033_            -----------SPCYHFGHKTRAKR-TLQRLCRS-VLSSSQ-SCTSYTGCEKLDDMKHYW 46
MYSL031_            ---------IFYPCYHFGYKTMAKR-PFICLCIYGSLKLSC-SGSSYTWMVKWVGWTDTL 49
MADH002_            --PPFSCLVRFIALLAFWISNKGQK-TPSEIMVGSDSLKLS-ELEQATQEVK-GWMTETL 55
MYDH019F_           --PPFSCLVRFIALLAFWISNKGQK-TPSEIMVGSDSLKLS-ELEQATQEVK-GWMTETL 55
MYSL030_            ----FLRTSKILALLVFWISNKGHC-TLQRLCR-SGSLKLS-ELEQATQEVK-GWMTETL 52
MYDH022X_           PLFLKTIVSFYSPPSILGYKHRAHKDPSGLYVAQGSSKALK-TLTYAPTGMSKNLGIPEK 59
MYDH055_            ----FYDLVRFYALLAFWISNKGQR-TLSEIMSDRVLSGFSGACASSTGMSKKLGCQKPL 55
MYLD08_             ----------------------FQK-NPSETMYDMVLSNS--QSCACYTGRYKVRMTETL 35
```

```
MADH011_            LVQNANPDCKSILKALR-----PGASLKEIMTACQGVGGPGHKARVLAK----------- 90
MYLDL03_            LVQNANPDCKSILKALR-----PGASLKEIMTACQGVGGPGHKARVLAK----------- 90
TKMH012_            LVQNANPDCKSILKALR-----TGATLEEMMTACQGVEGPGHKARVLG------------ 95
MYDH037_            WVQNANPDCKSIFESTRN----RGYYSKKCCQHAKECEDPAIKQGFSLE----------- 92
MADH005_            LVQNANPDCKSILRHSD-----QELHSEEMMTACQGVGGPSLKAGVWAE----------- 98
MYDH028_            LIQNANPDCKSILRALG-----PGATLEEMMTACQGVGGPGHKARVLAE----------- 90
```

```
MYLD07_             WVQNANPDCKAILSSLG-----AGASLEEMMTACQGVGGPSHKARILAC----------- 87
MYDH038_            LIQNANPDCKSILRALG-----TGATLEEMMTACQGVGGPGHKARVLAE----------- 87
MADH010_            LGPKCKSRLYVHLKSIRN----SFASLEEMMTACQGVGGPGHKARVLAE----------- 91
TLHC023_            TGPKCKSRLSVHFESIR-----TRGFIRRNDDSMPGSGRTRHKARVLAE----------- 67
MYSL033_            TGPECKSRLSVNFKSIR-----DRGYIRRNDDSMPGSXRTQPSTRVLAE----------- 92
TKMH010x_           ------------------------------------------------------------
MYDH063             K-HVLDPKCKS-----------------RLSVPFY--EHSGPVAFIRSK----------- 74
MYDH0035_           XXLYWIQKXQN-----------------PXWXXLFXKEHXGXXLLLWG------------ 77
MADH014_            CSKMQIQIVSPFSEHSDQ----GLHSKRCCQ-HAREWEDPAIKQGCWLR----------- 91
MYDH110_            CSKMQIQIVSPFYEHSDQ----GLHSKRCCQ-HAREWEDPAIKQGCWLR----------- 91
MADH004_            WSKMQIQIVNPFYEHSDQ----GLHSKKCCQ-HAKEWEDLAIRQGFWLR----------- 91
MYDH026-1_          WSKMQIQIASPFSEHSEQ----GLHYKKCCQ-HAKEWEDLAIKQGFWLR----------- 97
MYDH026-2_          WSKMQIQIASPFSEHSEQ----GLHYKKCCQ-HAKEWEDLAIKQGFWLR----------- 97
MYDH049_            WSKMQTQTVSPFCEHSDR----GLVRRIDDS-MPGSGRTLAIKQG--CS----------- 86
MYLDX005_           WSKMQIQIVSPFSEHSEQ----GLHSKKCCQ-HAREWEDHGHKARVLAE----------- 89
TLHC019_            CSKMQTQIVRPFYKHWKR----GFIRRNDDS-MPGSRRALAIMQGF-------------- 88
MYDH006_            WSKMQIQIVSPFYEPSDQ----ELHSRKCCQ-HVRESEDPAIKQGFWLR----------- 97
TLHC018_            WSKMHLQIVVSIL----------------------------------------------- 59
MYDH046_            WSKMQIQIASPFYEHSEQ----GLLLEEMIDSMPRSGRTQPSSKDFGLE----------- 95
MADH008_            LGPKXQX---------------------P------------------------------- 60
MYDH015_            LGPKCKSRLYVHFSRSIRFLTSLGEIEXPXXXXVVXEXPXLSXKGXSXLXAXSXXSXRLS 113
MADH003_            ------------------------------------------------------------
TLHC033_            S--KMQIQTVSPFSE----HSEQGLHSKKCCQHAREWEARPSSKSFGLR----------- 89
MYSL031_            LGPKCKSRLSVHFKS----IRTRGFIRKKCCQHAREWEARPYGKGFGLR----------- 94
MADH002_            LGPKCKSRIVRPF----------------------------------------------- 68
MYDH019F_           LGPKCKSRIVRPF----------------------------------------------- 68
MYSL030_            L-AKMQIQIVSPFY---------------------------------------------- 65
MYDH022X_           LLLGPRXAKIPXLXXAHFFX----XRLCXFXXXLSXXGGXXVSXXFXLG----------- 104
MYDH055_            IGPKLQNPELS------------GPFLRSFRFFGAHRXGXXXXDXFXRX----------- 92
MYLD08_             LGPKCXSRXGXPXIIXX------XIXXRXVXENXCQXXXLSXXTRXSKX----------- 78
```

```
MADH011_            --AISQVQGASIIMQKSNFRG---SEKDSVFQLWQKRTPSQKLQAPVGKRLCNCG- 140
MYLDL03_            --AISQVQGASIIMQKSNFRG---SEKDSVFQLWQKRTPSQKLQAPVGKRLCNCG- 140
TKMH012_            ---LRQ-------------------------------------------------- 98
MYDH037_            --AMSQVQHTNIMMQRGKLSEP---------------------------------- 112
MADH005_            --AIESSTAYKLNDAERQFSG--PEKISMFQTVAKK-------GPV---------- 133
MYDH028_            --AMSQGAKYKRNDAERQFLE--ARKELSVSTVAKKDTYPGIAGPLEKRA------ 136
MYLD07_             --GNEPSTKCKCNDCKEATLGGRPEKRLSVINCGKD-------GHSSQKFVG---- 130
MYDH038_            --AIESSTAYKHNDCREAFSG--PCKGLSASTCGKE-------GHLARNCRAP--- 129
MADH010_            --AMSQVQQTN-IMMQRATLR---VEKELSVSTVAKKDISPEIAGPYEKGLVEFWI 141
TLHC023_            --AMSQVQQHTNIMMPRGNFSR-AIKGLKRFNCGXEGHLARNL------------- 107
MYSL033_            --QCVKHKIQTYLMQ-RSNFRG-PIKGLRCFNCGKEGILARNCKVPRKKGLFGM-- 142
TKMH010x_           --------------------------------------------------------
MYDH063             --CSLA-------------------------------------------------- 78
MYDH0035_           --------------------------------------------------------
MADH014_            --QCVRYNIPTYCC-REVILEAKEGLSVSTVAKKGTSP---KIAGPYEKGCEN--- 138
MYDH110_            --QCVRYNIPTYCC-REVILEAKEGLSVSTVAKKGTSP---KIAGPSEKGCEN--- 138
MADH004_            --QCVKHTRQTYCC-RKAILGGQKKIKCFNCGKEGHLARNCRVLGKRAVKCG---- 140
MYDH026-1_          --QCARSAYKHN--------DAERQF------------------------------ 113
MYDH026-2_          --QCARSAYKHN--------DAERQF------------------------------ 113
MYDH049_            --ACAMSQVQNTY-----IIDAKRQ------------------------------- 104
MYLDX005_           --AMSQAQHTNIMM-QRGNFRGTEKDYVFQLWQRRTSSQKLQAPRKKAVK------ 136
TLHC019_            --------------------------------------------------------
MYDH006_            --QCGQAQLTNIMM-QRGNLEGQKRIKCFNWWQRRTCPQNWQAPSKKGC------- 143
TLHC018_            --------------------------------------------------------
MYDH046_            --AMSRAQQTNIMMMQRGNFKGQKRIKCFNCGKEGHLAR----------------- 132
MADH008_            --DCXX----LX--------XX-XXXFFXFYXX---GG-------XRX-------- 83
MYDH015_            KLXCSXXXXXFX--------GAXSKEFXXFXXGXXXXGXLWPXNXLRPXEXGXX-- 159
MADH003_            --------------------------------------------------------
TLHC033_            --QCVKYN----------------IQTYCCMR-GNFRG------------------ 108
MYSL031_            --QCVKSADKRNDAERQFSGDKKGIKCFNCGKEGHLAGTAGPLEKRLWK------- 141
MADH002_            --SKLRDQ------------GRHWEGMMDSMPG----------------------- 87
MYDH019F_           --SKLRDQ------------GRHWEGMMDSMPG----------------------- 87
MYSL030_            --EQLFXX------------AXXGXKWMDSMSXXXXKPPXLKSQ------------ 95
MYDH022X_           --------------------------------------------------------
MYDH055_            --XGGRPXAXK--------------------------------------------- 101
MYLD08_             --GXSACXTLX--------------XYXHXXXRXXDRH------------------ 100
```

c) Gag subtype C alignment

```
MADH009_         ---------GVSPAAFWKYDKGQRNPLET-MSDRFFKTLR-AEQATQE-VKNWMTETLLV 48
MYDH111_         ---------GVSPAAFWKYDKGQRNPLET-MSDRFFKTLR-AEQATQE-VKNWMTETLLV 48
MYDH004_         ----------SALAAFWTSDKVQKNPLET-MYDRFFKTLR-AEQATQD-VKNWMTETLLV 47
MYDH016_         ---FVNNSSDVSPAAFWTSNKGQKNPLET-MSDRFFKTLR-AEQATQD-VKNWMTDTLLV 54
TKMH017_         --------------QHFGHKTRAKGPFRD-YVDRFFKTLR-AEQATQE-VKNWMTDTLLV 43
MYDH007_         ------------ALQYFGHKTRAKDHLRD-YVDRFFKTLR-AEQATQD-VKNWMTDTLLV 45
MYDH0020_        -GFLMEQSGCLALSAFWDIKQGPKEPFRRLWYDRFFKTLRSWNKLHKM-SKIGWTDTLLV 58
MYDH024_         ------------------------DYVR-----PVLSNLKSCTSYTSM-YKIWMTDTLLV 30
MADH015_         ---------RLYSPAAFWTSDKDQRNLLETMSTGSLK-LYELNKLHKM-SKIGCQTPCWS 49
TLHC120_         ---------RLYSPAAFWTSDKDQRNLLETMSTGSLK-LSELNKLHKM-SKIGCQTPCWS 49
TLHC012_         ----------FYRPSSILDIKQGPKDPLETMSTGSLK-FYELNKLHKK-SKIGCQTPCWS 48
TKMH010_         ---------SIYSPAAFWTSNKGQRTFYGLCRPVLSN-IIELNKLHKR-SKIGCQTPCSV 49
MYDH040_         ----LFKASEVYSPAAFWTSNKGQRSPLETMSTGSLK-PSELNKLHKM-SKIGCQKPCYV 54
MYDH012_         ------------GPAAFWHKTRSKGTFSRLCRSVLSN-FKQLNKLHKM-SKIGCQTPCSV 46
MYDH039_         -------LAFCVAPQHFGHKTRTKRTLSRLCTTDSLQ-LSELNKLHKM-SKIGCQTPCSV 51
MYDH018_         -----------LALQHFGHKTRAKGPFRDYVCPVLSN-LKSCTSYTRC-QKLDGQTPCWS 47
MYDH0021_        -------NSKMLALQHFGHKTRAKEPFRDYVGPVLSN-LKSCTSFSRC-KKLDGQTPCWV 51
MYSL1030_        -----------RPRQHFGHKTRAKEPFRDYVDRVLSN-LKSLNKLLKM-YKIGCQTPCWV 47
MYDH013_         ------PSKFVYPRSILDIKQGPKEPFRDYVRPVLYN-FKSGTSYTRC-KKLDDRHLVG- 51
MYDH062_         --------RGLWTLQHFGHKTRAIGTLSKTSGXGSLKPSRSCTSYTXMXKNWDDXKPWXV 52
MYDH002_         ----FYSLVRFIALQHFGHKTRTKGTLSRLCRTGSFKP---------------------- 34
MYDH023_         ------------APQHFGHKSGSKVTLXXLSYDRFFKTLSSCTSYTRX-KKLDDXTPCAV 47
MYDH036_         ----------FSPSSILDIKQGPKDPLEIMWT-GFSN-LKELEQAFQDVKKLGCETPLLG 48
MYDH027_         VKQGVLKTGRFYPFIIFDMKQWPNGPFIYYGRPGLSN-FKDLEQSTQDGKNLMNMTPFS- 58
MYDH047_         ------------GPAAFWTSNXXQXNPLXTMSTGSLKLSE-QNKLHKI-VKNWMTXHLVG 46
MYDH052_         -----FGAYYFGGPVAFWTSNKVQKNLLETMSDRFFKTLR-AEQATQD-VKNWMTDTLFG 53
MYDH009_         ------------SPCSILDIKQGPKDPLETMYDRFFKTLR-AEQATQD-VKNWMTGYLVG 46
MYDH0056_        -----LLGQCDFSPAAFWTSDKAKGPFRD-YVDRFFKTLRAEQATQEV--KNWDDRHLVG 52
MYDH001_         ----FLINRRMYSPAAFWT----------------------------------------- 15
MYDH011_         ----------------FWGEKPAXXP---------HLKLSSGQATQDV--KNWIDRHLV- 32
MYDH048_         ----PPFEHSRFSPPAFCQRQG-LGPFRD-----LGPVLKTQSCTASQDVKNWIQTLLGX 50
MYDH003_         ---------WILPLQHFGNKTRPNYPLET-MSDRFFKTLRSWNKLHRE-VKIWMTETLVG 49


MADH009_         QN-ANPDCKT-ILRALG-PGATLEEMMTSCQ-GVGRPGHKARVLAEAMSQMNN---ANVM 101
MYDH111_         QN-ANPDCKT-ILRALG-PGATLEEMMTSCQ-GVGRPGHKARVLAEAMSQMNN---ANVM 101
MYDH004_         QN-ANPDCKT-ILRALG-PALHSKRCCQHAR-EWEDPVIKQEFWLEAMSQATNNVNAAIM 103
MYDH016_         QN-AEPRLYD-HFKSIR-ARGYFRRNDDSMS-RSGRPGHKARVLAEAMSQANS----SIM 106
TKMH017_         QN-ANPDCKS-ILRALG-PGATLEEMMTACQ-GVGGPSHKARVLAEAMSQANQ---TSIM 96
MYDH007_         QN-ANPDCKT-ILRALG-PGAYIRRNDDSMS-GSWAASPQS--MSVACGNEPKQIIQTSC 99
MYDH0020_        QN-ANPDCKT-ILRALG-PGATLEEMMQHVR-EWEGLPQSKGFG-RALTKSTI---HTYC 110
MYDH024_         QN-ANPDCKT-ILKALG-TGATLEEMMTACQGEWEDLGTKQGFWLRQCAKETM---PTSC 84
MADH015_         KM-LNPDCKT-ILKALG-TGATIRRNDDSMSGSGRPGHKARVL-AEAMSQLNN---TTVM 102
TLHC120_         KM-LNPDCKT-ILKALG-TGATIRRNDDSMSGSGRPGHKARVL-AEAMSQLNN---TTVM 102
TLHC012_         KM-QNPDCKT-ILRGFG-QGATL---EEMMTAC--------------------------- 75
TKMH010_         QN-ANPDCKS-ILRALG-PGATLVRNDDSMSRSGRTSPQSKSVGLRLMSQANQ---TSII 103
MYDH040_         QN-AEPRLQD-HFKSIR-ASGHIRGNDDSMSGEWEASPQSKSV-VCAMSQANN---TNIM 107
MYDH012_         QN-ANPDCKT-ILRALG-PGASLEEMMTACQRVGGPGHKARVL-AEAMSQVNN---SNIM 99
MYDH039_         QN-AEPRLYD-YFKSIG-TRGYIRRNDDSMQGVGGPGHKARGL-AEALSQVNN---PHIM 104
MYDH018_         KM-QT-QIAR-HFKSIRSRGY-IRRNDDSMSRESEAWPQSKECWLRQCAKQTM---RTYY 100
MYDH0021_        QN-AT-QIVR-PFYEHYGQGHPLEEMMTACQGVGRTQPQSKSAWLEELSQATN---ASYC 105
MYSL1030_        QN-AEPRLYD-HFKSIRARGL-IRRDDDSMSGSGRTQPQSK-SLLRQLSQATN---ASIM 100
MYDH013_         PK-CEPRLYD-HFKSIRTRGFIRRND-DSMSGSGEDLGHKARVLAEAMSQGNN---TYIM 105
MYDH062_         QKXXNPGLYE-PFXEAXXFFFXXXWGXXCXQXXKVXXXGPLXXXARXXXXXXX---XQXX 108
MYDH002_         ------------------------------------------------------------
MYDH023_         QN-ANPDLSE-HFKXX---------------------LGPXGXXSEVNRWT--------- 75
MYDH036_         PKMRNPDWKDHFLRAL---CFXXXX---------XGGNYRIGIN---------------- 80
MYDH027_         AKMRNPDWKT-ILRALGPWGFIRKKGLDSLA---RGVGRDLAIKARSF------------ 102
MYDH047_         PK-CEPRLSD-HFKSIRTRGGGSKKCCQHVREWEDLATKQGFL----------------A 88
MYDH052_         PK-CEPRLSD-HFKSFRTRGYTRKKCCQHVREWEDPATRQELLGLRQLSQINNA---NII 108
MYDH009_         PK-CEPRLSD-HFKSIRTRGFIRRNDDSMSGSRRTWPQSKGFGCGNEPSNQYN------H 98
MYDH0056_        SKMRNPDLSRLFLRSIRASGPTLWGNCCPALSRGLGRPLANKAKSVPLEGNGTRQTNSNL 112
MYDH001_         ------------------------------------------------------------
MYDH011_         ---------------VPKCEPRLQEHSYVALGPGAG---------------GSKEYDRL- 61
MYDH048_         KX-QTXDWQT-LFSDXXXFWPTXXGKCSPLXXGXWXRPLAXXQRXPXXSXEPX------- 101
MYDH003_         PKCCXQVWXX-LFXXPX------XXLXXPXGGGXXTTSXSRGVXXXLATXQGX------- 95
                             :

MADH009_         MQRNNFKG-PRRIVKCFNCGKEGTSPKIAGS-RKKGVE----- 137
```

```
MYDH111_            MQRNNFKG-PRRIVKCFNCGKEGTSPKIAGS-RKKGVE----- 137
MYDH004_            MQRGNFKG-PKRSVKCLTCGKEG---HIARN-CK--------- 132
MYDH016_            MQKSNFKG-TKRTFKCFNCGKEGTSPEIAGPPRKKGFG----- 143
TKMH017_            MQKSNFKG-PKRNIKLFQLWQGRASSQKLQAPSEKGLC----- 133
MYDH007_            CRRGNFYGSPRGTIKCS-------------------------- 116
MYDH0020_           CRETILKA--LREWLNVQLWQGRALAQKLQGPSE--------- 142
MYDH024_            C------------------------------------------ 85
MADH015_            MQKSNFKG-PKRLLNVSTVARKAPSSK-LQGPRKKGGGN---- 139
TLHC120_            MQKSNFKG-PKRLLNVSTVARKAPSSK-LQGPRKKGGGN---- 139
TLHC012_            -------------------------------------------
TKMH010_            DASKQFSG-PYKIYSMFQLWQGMASSQEIAG------------ 133
MYDH040_            MQKSNFKG-PRRPVKCFNCGKEGPISQEIAGPSEKGLLEMWG- 148
MYDH012_            MQRNNFSR-PYENYYMFQLAARKGTSPEIAGPLGKRAFEN--- 138
MYDH039_            MQ----------------------------------------- 106
MYDH018_            CREAFLKG-LKEMLSGSTVAKEGTSQEIAGPLGKGLLERIGIN 142
MYDH0021_           CRKAISRA-LKELSNVQFVARRGNW------------------ 129
MYSL1030_           MQ----------------------------------------- 102
L10_                MQKSNFKG-PSKNYSMFQLWKGRTFSQKLQAPRKKAGG----- 127
MYDH013_            MQRNNIKG-P--------------------------------- 114
MYDH062_            XSX---------------------------------------- 111
MYDH002_            -------------------------------------------
MYDH023_            -------------------------------------------
MYDH036_            -------------------------------------------
MYDH027_            -------------------------------------------
MYDH047_            DA----------------------------------------- 90
MYDH052_            DACKSIFSGPSKNYSMLQLVAWEG------------------- 132
MYDH009_            NDAKKQFSRPYENCYMFQLVAREGTSPEIAGPSEKGL------ 135
MYDH0056_           NGAEKRIFSGSSENLLK-------------------------- 129
MYDH001_            -------------------------------------------
MYDH011_            -------------------------------------------
MYDH048_            -QTPHLM------------------------------------ 107
MYDH003_            XPXAK-------------------------------------- 100
```

3. *GaG* ClustalW alignment

Amino acid sequence from *gp 120* subtypes were aligned according to subtypes (A1, C and D) to determine if they were conserved. The areas marked in red colour indicate conserved regions. While blank area indicate regions with high variability in the amino acid sequences.

a) gp 120 subtype D alignment

```
MADH012_          FLGVKPVVQLNYCWNGSLAEEDIVIRNCKSHKSCKKHNSTAS-CVCNNSLHKALQQYKTR
MYDH056_          ----KPVVSTQLLLNGSLAEEDIVIRNCKISQIIAKNIIVQLNESVTINCTRPYNNTRQG
TLHC009_          ---------TQLLFNGSFAEEDIIIRSENLTNNARTIIVHLH-KTLLLNCTRPNNNTRQG
TLHC021F_         ---------TQLLFNGSFAEEDIIIRSENLTNNARTILVHLH-KTFMLNCTRPNNNTRQG
MYDH057_          FLGVKPVVQLNYCWNGSLAEEDIVIRNCNLTNNAKNIIVQLN-ESVTINCTRPYNNTRQG
TLHC102_          ----------PLLLNGSLAEEDIIIRTANLTDNTQNIIVHLT-ATVMINCTRPNNNTRQG
TLHC040_          ----RSANSAINCCEWQSSRKRDNNSICKYHKQCQNYNSTSY-ESVTINCTRPNNNTRQS
MYDH062_          ---INQWCNSIARCMAVSQKKTYYLGSENLTNNAKTIIVHLN-KTVMINCTRPNNNTRQG
MADH007_          ---INQWCNSIARCMAVSQKKTYSLGSENLTNNAKTIIVHLN-KTVMINCTRPNNNTRQG
TKMH015_          ---------LHCCKWQSSRRRDNNSISKISQVILKTSYVQFN-ESITINCTRPNNNTRDG
TLHC023_          ----RSANSAINCCCMAVSQKKRSSLDLKISQTMPKLSSYILMSLSQLIVPGP-TTIQGK
```

```
MADH012_          YAYRTRASTLYNKDNRRYKTSTLYHSWSSMESNITTGSSKIKGPFYPDKSNLSSTRGRGT
MYDH056_          THIGTRASTLYNKDNRGYKTSTLYHSWTAWNTTLQQGSSKIKGAFSPDKSNLSSTRGRG-
TLHC009_          VHIGPGQAIYTTDVIGDIRKAHCNISRADWNKTLQHVAMQFRNLLHLPRIHFNSSSGGDP
TLHC021F_         VHIGPGQAIYTTDVIGDIRKAHCNISRADWNKTLQHVAMQLRNLLNLPRINFNSSSGGDP
MYDH057_          THIGPGQALYTTKITGDIRQAHCNISGAAWNKTLQQVAKKLRGLFNLTKVIFHPPAGGGP
TLHC102_          VHIGPGQALFTTEVIGDIRKAYCTIRTADWE-----------------------------
TLHC040_          VRIGPGQAFYTTNIIGDIRQAHCNISGTIWKKTLQRVANKLREHFN-KTINFTQSSGGDP
```

```
MYDH062_       VHLGPGQALYTTEVIGDIRQAHCNISREDWNRTLQQVAIKLGKPFYPDNNNFYAILRRGP
MADH007_       VHLGPGQALYTTEVIGDIRQAHCNISREDWNRTLQQVAIKLGKPFYPDNNTF--------
TKMH015_       VHIGPGRAYFTTEIVGDIRKAYCTINETQWGKTLYQVASQLGVFLTGTNITFYPTCGRGP
TLHC023_       VCVSDQGKHSIQQTSSETYDKHIVILVEQYGRKLYNGSLTNSENTLTKQSILLNPREGTQ
```

```
MADH012_       QKLQCTVLIVEGIFSGIQQSGYSSWGGECYSPIVLESITSMRVNNFNM------------
MYDH056_       SRISMHALIVEGNFPLHSQICS--------------------------------------
TLHC009_       EITTHSFNCGGEFFYCNPHDCFIAHGQMAL------------------------------
TLHC021F_      EITTHSFNCGGEFFYCNPHDCFIAHGQMAL------------------------------
MYDH057_       RNYNAQFYLWRGFSLEYNNLVNSHGEESVIPQSYSSPSPPCGYTIFT-------------
TLHC102_       ------------------------------------------------------------
TLHC040_       EITTHSFNCGGEFFFCNTSGPVYSSWNALTARG----------TCVTYTSPCRIKQLYHV
MYDH062_       EIQHTALIVEGNFSTAIPHNCLIAHGKWDRGYSSNACHHLHADKQFYMWRRGSM------
MADH007_       ------------------------------------------------------------
TKMH015_       LENYKHTALIGRRV----------------------------------------------
TLHC023_       KLQHTALIVEGSFSSAIHQGLFNSLGMHYQPGEHECHIILHAESNNYNM-----------
```

```
MADH012_       ------
MYDH056_       ------
TLHC009_       ------
```

b)gp 120 Subtype C Alignment

```
TLHC001_       ---------NSTTAEWYVSLEQEIIITI--------------------------------
MYDH058_       -----SQWYQLNYCSWYVYAEEEIVIRISNLTNNAKTIIV--HLNESVEIVCTRTRSIIQ
MYDH035_       ---SKPVVSTQLLLNGSLA-EEEIVIRISKSDRQCQNNNS--TASRPCKHCVVQGPNNNT
MYDH017_       ---DSGQWYQLNYCYGSLA-EDETIIRICICTN-NAIIIV--QLNKTVAINCTR-TQQYT
MADH015_       ---DNQWYQLNYSLNGTPSRKRDNNSISKICQTMPKQSIV--HLKTPVEIECIR-PGNNT
MYDH032_       ---DNQWYQLNYSLNGTPSRKRDNNSISKICQTMPKQSIV--HLKTPVEIECIR-PGNNT
MADH008_       -----SPVVSTQLLLMVVSQKKTSYLGSENLTNNAKTIIV--HLKDPVEIVCTRPNNNTR
MYDH009_       -----YPVVSTQLLLMVVSQKKTSYLGSENLTNNAKTIIV--HLKDPVEIVCTRPNNNTR
TLHC012F_      ---------STQLLLMVVSQKKRSSLGSENLTDNVKTIIV--HLQNPVEIVCTRPNNNTR
TLHC013_       ---------STQLLLMVVSQKKRSYLGSENLTDNVKTIIV--HLQNPVEIVCTRPNNNTR
MYDH031_       ------NRVSTHYCSMVVCRRKDS-NSSENLTNNVKLIIV--HLKKSVDIVCTRPNNNTR
MYDH036_       -----KPVVSTQLLSWSSRRRGDRNSGSKNLTDNAKIIIV--HLQDYVEIVCTRTRQYYK
TKMH017_       ----------SLLLNGSLAEEEQIIRS-ENLTDNTKTIIV--HLNESCRNSLYK-AQQYY
MYDH045X_      FFEMKPVVSTQLLLNGSLAEEDIVIRISKSGTDNAKIIIV--HLNESVRIVCTR-PGNNT
MADH002_       ----YASSINSTTVKWSSSRRRDSNSDLKICQTMPKYSSY--SLKT-LYTLCGTRAQQSY
MADH013_       -----KPVVSTQLLSWSSSRRGDSNSDLKICQTMPKSSSY--IFKT-MSRLCVQGPGNNT
TKMH018F_      -----SAICINSTTVNGSLAEEKIVIRISRSDRQCQNNNS--TSSRTLSKLCAQGPAIIL
TLHC012_       --------FQLNYCSMVRLAEERSYLGSKDLDRQCQNNNS--TSSRPCRNCVYTGPAIIL
MYDH019F_      ----INQWYQLSYYSNGSIAEDEIIIRIWKICKQCQNNNS--AASCTCRNKLYQDLTIIQ
MYDH013_       -------RINPLLLNGSLAEERGSNSSRRSWTNNRQQSMIGDILTESVRNCVYKAQQSNT
TLHC014_       --------STQLLLNGTFTQRETYSLDLKICQTMPKYYSY---SLMNLSQLIVCDPTIIQ
TKMH011F_      --------STQLLLNGTSSSNKRYYLRSENLTDNAKIMIV--QLKDPVEIECTETQQSYK
MYDH044_       ---IKPVVSTQLLLNGTSSRRRGGNYGLENLTNNAKTIIV--QLKKPSRNSLLQMQPSHF
MADH001_       ---RLSQCINSTTACMVVSQKKRSSLGSENLTNNAKTIIVTSYNPCSTVLLCITLRRGPR
MADH011_       --------------------CRDNNPDLKMCQTSAKTIIA--HLKEPVTNTSERDLAINT
MYDH2663_      ----AKPVYQLNYCMHGRLAEEEIVIRICKSAKQCQNNNSHILQPLSYCPSLHHTQEGTQ
MADH009_       -------RRAESTDSICEHYKQWPKTYSSNFTKHARNYVG----------TQTLTTNTST
MYDH018_       -------WGPSYLPAMFTICFILHGSVMVAFAVLLVVLIALLHLVYTCQQIRLYEQYAKF
TLHC018F_      ------LFGARLLLHMSIICFILHGSVDGFICITISTISPRMYHSTAMDLLRLLKMLLPQ
MYDH011_       ---------PCHVYNLFYSAWECRWYALVSLPVLLFSIWYY-YTRKDCYLYNKKPPATDK
MYDH043_       ---------FWGLPPLPFLSFVLFCIGVSWLHLKYYSYYSICMYCSTVLMYYNRKNSPPQ
MYDH010_       ----LLGARLPCTCLSFVLFCWECDGCISIIVCTIVSYYSTCMYYSTVWMCCNKKNSPLQ
TLHC024F_      ------CPFLSFVLFCIEVCWLNLKYYSYCYDMYVLLNSFGCIAICKILLHNSSYVLYSL
```

```
TLHC014F_       ---------LSLFAVFCIEELYVCINPSRSCETCVVLILWTPSAMLSDRFICYEDYSFTH
MYDH016_        -------PTQLSWYGSTVSRAAERIFPYQDWLMSWCFVLVDYSNCQGRVQGTFQSMCVTY
```

```
TLHC001_        ------------------------------------------------------------
MYDH058_        GKSIENRTRTNFLCNRRLSMGVKTGPLHIGVGELGLKLFHRE-VGNSKTFPSIN---IKF
MYDH035_        RKSIRIGPGQTFYATGDIIGDIRQAHCNISN--WNQTLQNVT-KQLQKLFNKTI---AIY
MYDH017_        RKSMRIGPGQTFYATGDIIGDIRQAHCNISGEEWNETLEGVK-EKLKEHFPVET---SHL
MADH015_        RRSVRIGPGQTFFATGDIIGDIRKAHCNINGSKWNETLREVS-EKLREHFPNTTNHNMYA
MYDH032_        RRSVRIGPGQTFFATGDIIGDIRKAHCNINGSKWNETLREVS-EKLREHFPNTTNHNMYA
MADH008_        -KSMRIGPGQTFYATGEIIGEIRQAHCNISEEKWNETLQGVG-RKLQEKFPNST---IKF
MYDH009_        -KSMRIGPGQTFYATGEIIGEIRQAHCNISEEKWNETLQGVG-RKLQEKFPNST---IKF
TLHC012F_       -KSMRIGPGQTFYATGAITGDIRQAHCNISEKAWEDTLQKVG-GELHKHFPNKT---ISF
TLHC013_        -KSMRIGPGQTFYATGAITGDIRQAHCNISEKAWEDTLQKVG-GELHKHFPNKT---ISF
MYDH031_        -KSMRIGPGQTFYATRDIIGDIRQAHCNISEQEWNNTLQEVV-TELHKHFPNKI---IRF
MYDH036_        EKSLKIGPGQTFYATGAIIGDIRQAHCNISEKKCECHFTKKS-VENYKNTSLIK---QYA
TKMH017_        KSK---------------------------------------------------------
MYDH045X_       RQVCGSDQDKHSMQQETWGGNLRQAHCNITDENGMPPSEGWK-KSKPSYST---------
MADH002_        KEKYKDRTRTNILCNRRHNRRYKTSTLSHQ------------------------------
MADH013_        RKSIEDRTRTNILCNRSHNRRYKTSTLSHQCKEMRMTLYKEVGRKLQKHFPNKT---ISF
TKMH018F_       GKVCGSDQDKHSMQQETYS-EIYDKHIVTSVKRNGMLLYKRS-VKNYTNTSLVVKQSAFK
TLHC012_        GKVWRIGQDKHSMQQETSS-EI--------------------------------------
MYDH019F_       EKVSGIGTRDKHSMQQGALVEIQDRPIGHSGRNGINLSKAGKIEELSLIDHYLLT-----
MYDH013_        RKSIRLGPGQTFYATGDIIGNIREAHCNISEVQWNKTLKGSLKNYETLPSYNYDLISTIR
TLHC014_        EKVSESDQDKHSMQLGDIIGDIRQAHCNISGENWTKTLERVK-EKLKKHFSNNK--TIQF
TKMH011F_       RKYKDRTRTDTLCNRCHNRGYKTSTLYCQYIKMESNFTRGRCKISKALPYSNNK----IY
MYDH044_        NAESNKLYECGRGSAGNVCPSLPR------------------------------------
MADH001_        NYNTYLSLWRRIFLLQVHQDCLMEHTWLIVLQILLQMRPSHSNAESNNLSTCGKG-----
MADH011_        SNKHKDRTMDQTLLSNQGVHSGGGYQDKHYLYLY--------------------------
MYDH2663_       KLQHIPLLVEENFSTASTSRLFNGTYMANSTSNSTSNATITLQCRIKQFIHMWQ------
MADH009_        KGFH--------------------------------------------------------
MYDH018_        SSTH--------------------------------------------------------
TLHC018F_       EKAHAAEITSSPAVIEARAHSYGWMPKECFKLNAYLSSSSILFCSTDPSMMHGSYTCSGY
MYDH011_        AMPLRYARSTPDDGSDFSVIVSGSVPVVLHLPLQGYSKLLCTNDSNLLGLSAAIN-----
MYDH043_        LKLCVVISRSPPEDGANFIVLLGKCPCNFFPTSCKVLFQSLSLMVQCACLISPYYASCCI
MYDH010_        LKLCVVISRSPPGGWCKSYLCYVRVVFSSFFSYLFLKFIPLSLNSCYMVLLYYFLIIPPV
TLHC024F_       GPLLRMVSRLLFHSGSAFVNCLPTSWSSVRSMPFHSYYMCFGLESAYHGFLAASN-----
TLHC014F_       LQAMPLQSPRVLATVCCFNTLLCLYTVSMRLSNVLLTSCRAFSYKSYPYQQHGLS-----
MYDH016_        LHTARACTLGIGMRTSNVCHSHSKEVKLPPAENRLSLTLSSNITMPLIL-----------
```

```
TLHC001_        ------------------------------------------------------------
MYDH058_        APSSGGALDLERLSLNW-------------------------------------------
MYDH035_        NLPQGGDLEITTIALIVGGEFFYGNTSDLLNGP---------------------------
MYDH017_        HQPQEG-PRNYTHSFNCRGEFFYCNHQDCLIVWVLEIHLTSMRINNFTVQRRTHDPP---
MADH015_        PSSGGATSIFTLHSFNCRGEFSVHYIQLIKQHIHVIMAKHTHTMYRHVLTHPLRQLSIYM
MYDH032_        PSSGGATSIFTLHSFNCRGEFSVHYIQLIKQHIHVIMAKHTHTMYRHVLTHPLRQLSIYM
MADH008_        APSSGGASKSQHIALIVEENFSIAIHQNCLIVHTCYSYYWSSIQPHTPCRINNYNMWQGE
MYDH009_        APSSGGASKSQHIALIVEENFSIAIHQNCFN--STYMLIVLIVLNSTSHSMQDKQLYHVA
TLHC012F_       KPSSGGASKFN-TSLSLSRRIFLLQSSKLFNSHTCLIVIHANSTNRPKFT----------
TLHC013_        KPSSGGASKFN-TSLSLSRRIFLLQSSKLFNSHTCLIVIHANSTNRPKFT----------
MYDH031_        APSSGGDLEIQHIALIVEENFSIAIHQNCLIVHTFSMLKCYAQMLINLTLLQIKQLYHVA
MYDH036_        LHPPQEGPRTYNTSLYLSKRIFLLHTSKLVNRSYNRPTCLTDLSSQII------------
TKMH017_        ------------------------------------------------------------
MYDH045X_       ------------------------------------------------------------
MADH002_        ------------------------------------------------------------
MADH013_        APSSGGASNLQHIALIVEENFSIAYIKTGYSVIYSSYMSNRLIVSNH-------------
TKMH018F_       QAIPRRGPRNYNTSLYLSRRFFYCNTSELFKESYMVNGTDMLTMDLNSTLRLMPIKQIY-
TLHC012_        ------------------------------------------------------------
MYDH019F_       --YPRGGPSKFTPWLNWGGKIFFWQHKPG-------------------------------
MYDH013_        GGLLRTTPRSFSGGKNSMCTHPAASSSSCLVVLSSADILLINIS----------------
TLHC014_        ANSSGGDPEIQHIALIVKESFSIAIQQYCLIIISQMKPHNP-------------------
TKMH011F_       CILRRGPRNYNTSLYLSGENFSIAIHQSCLT-----------------------------
```

```
MYDH044_        ------------------------------------------------------------
MADH001_        ----VGRAMYALHSR---------------------------------------------
MADH011_        ------------------------------------------------------------
MYDH2663_       ----GGRTSNVCPSFKG-------------------------------------------
MADH009_        ------------------------------------------------------------
MYDH018_        ------------------------------------------------------------
TLHC018F_       MPLKKDLCWPSSTWPCTNTWAWTNCM----------------------------------
MYDH011_        ----GSCCYKRQPWPY--------------------------------------------
MYDH043_        EWLPGPILIPSLDYNGPWPYNWTCLKMTNIIWELAC------------------------
MYDH010_        ASKWLAWALSLSFSWING------------------------------------------
TLHC024F_       ---GLAWALSSIPTPGI-------------------------------------------
TLHC014F_       ------NTLQSYPWYKYRC-----------------------------------------
MYDH016_        ------------------------------------------------------------
```

c) gp 120 subtype A1 alignment

```
TLHC021F_       ------------------------------------ISQYYQLN--CCKWQSSRRRDKDS 22
TLHC004_        ------------------------------------ISQSYQLN--CCKWQSRRRRDKDS 22
MADH006_        -------------------------------------DQACSINSTAVEWHSSKRQGNNS 23
TLHC018F_       ------------------------------------------FNSTTVEWQFSRRRGKNS 18
TLHC007_        ------------------------------------------FNSTTVEWQFSRRRGKNS 18
TLHC037_        ------------------------------------FGTSQSFNTAVVKWQPSRKRGNNS 24
TKMH002_        -------------------------------------------------SMAVSQKERSC 11
MYDH036_        ------------------------------------------------------------
TLHC005_        ------------------------------------------VQLNCCKWQSSRRE-GKN 17
MYDH004_        ----------------------------------------AVVQLNCCYNGSLAKEKVKI 20
MYDH037_        ---------------------------------------IKPVVSTQLLYGSLAENQTKI 21
TLHC016F_       ---------------------------------------SQSFNSIAVKWQSSRK-GSNN 20
TLHC002_        ---------------------------------------SQSFNSIAVKWQSSRK-GSNN 20
MYSL032_        ---------------------------------------KPVVSTQLLLNGTLAKDRGNN 21
MYDH063_        -----------------------------------FFRLGQSVSTQLLLNGSLAKEEVKI 25
MADH010_        --------------------------------------TRQRASTQPLLNGSLAYEQLII 22
MYDHX5_A1       ------------------------------------------------------------
MYDH050_        --------------------------------------TSQCA-TQLLLNGSLAKGEVKI 21
TLHC007F_       ----------------------------------------ASSSTQLLLNGSLAEGKVII 20
TLHC017_        ----------------------------------------ASSSTQLLLNGSLAEGKVII 20
TLHC012F_       ----------------------------------------ASRSTQLLLNGSLAEGKVMI 20
TLHC033_        --------------------------------------------PIAVEWQSSRRRGNNS 16
MYDH046_        --------------------------------------KPVVSTQLLLNGTSSRRRDNNQ 22
MYDH039_        ----------------------------------FFGKRPVSINSAAAKWQSGQKIRQKL 26
L12_            ----------------------------------------------LCCWQLAEERNNS- 13
L25_            ------------------------------------------NSLLSMAVYQ---KERSC 15
TKMH019X_       --------------------------------------------LLSMAVYQ---QMGQK 13
MADH014_        ------------------------------------FLXTGPQYQLNYCKWQSRRRRGKN 24
MYDH008_        ---------------------------------------QASSINSTAVKWQSSKRRCNN 21
TLHC010F_       --------------------------------------CDQASSATNCCCMAVSQKKRSY 22
TLHC011F_       ----------------------------------------SQSFNSIRCQMAVQQKKRQK 20
TLHC024F_       ---------------------------------------TSQYFNSTAVEWQSSRRRSSN 21
TLHC036_        ---------------------------------------TSQYFNSTAVEWQSSRRRSSN 21
TKMH042F_       -------------------------------------FGQASSSTHCCSMESSRK-RSKN 22
TLHC003_        -------------------------------------FGQASSSTHCCYMESSRK-RSKN 22
TLHC011_        ----------------------------------------ASFSTQLLLNGSLAEGEIRI 20
MADH005_        ----------------------------------------AVVQLNCCYMAVSQKKRSRL 20
L26_            ---------------------------------------------LHCCKWQSSRRRGKD 15
TLHC0024_       ---------------------------------------PASSSTQLLVKWQPSRKRGKN 21
TKMH017F_       ----------------------------------------------TVVKWQPSRRRGNN 14
TLHC014F_       -----------------------------------FLGPASSSTQLLLNGSL-AEKEVII 24
MYDH027_        -----------------------------------FLGQARSINSTTVNGSL-AEGEVRI 24
MYDH033_        -------------------------------------DQACRINSTAVAMALSAKDKVII 23
L15_            ----------------------------------GFLGASAFCXXSTAAKWSSSXKGNSN 26
TLHC039X_       ---------------------------------------SAFCSTPLLLNGSLAEGKVKI 21
TLHC018_        ----------------------------------------SQSFNSIRCKWQSSRRKDKN 20
```

```
MYDH021         ------------------------------------------------------------
MADH004_        -------------------------------------------FQLNHCYMAVSQMKRND 17
MADH003_        ----------------------------------------RSCKYHKQWS------KDNS 14
MYDH026X_       --------------------------------------LLCSCAWLVRTQLRVKGCRDDV 22
MYDH040F_       ---------------------------------------WALPISAHILNCFILHRRIIA 21
MYDHX4_         ------------------------------------------------------------
TLHC023_        --------------------------------------FLGSSCFCHIHNLLYSARESYS 22
MYDH053_        -------------------------------------LFGDLPTFCHIYNVALFCLEFYT 23
MYDH019_        ------------------------------------PPVFAISIICLILHLRVIIVLFAS 24
MYDH025_        -------------------------------------LFGLSPAFCDIHNLLYSAWESYS 23
MYDH021_        --------------------------------LFGLVLSLPIFIICFILHWRVILPSVLP 28
MYSL033_        -----------------------------------CFLCISQFALVCMGELYCPCYLILA 25
MYDH042_        -------------------------------------------GFLRCIAVVYLN--AVN 15
MYDH028_        ------------------------------------STPRKSVGYRDQGPSVLMQQGNII 24
TLHC006_        ----------------------------------FWG-----XALX-GXSRGAX--P---- 15
MYDH045_        ---XXICXX---DX-AXVXLQLXVXVXCRX-XXLXGXLVSDWPIPPAFCNXLLXLPXRDK 52
MYDH015_        -------------------------------YAYRS-----SPSEPQQKHLILFRPYNRA 24
MYDH050F_       GLHALLCPAXTLNPXXGPXKXXLSHXHXRLXXRILDSLPQPKPLLYGRXXXLXDMLSSXA 60
MYDH038_        ----------------------------------------FWLRALLCYSSNTSSFLDPS 20
MYDH038_A1      ----------------------------------------FWLRALLCYSSNTSSFLDPS 20
MYDHX3_         ------------------------------------------------------------
MYDH024_        ----------------------------------------FXXLSLXXIXXSTAAADXKG 20
MYDH052_        ----------------------------------------------APQXLPXQXXSSXP 14
MYDH014_        ------------------------------------------------SXXPAVX-EGXX 11
MYDH049_        -----------------------------------------------SLQGNPALCEGCV 13
MYDH002_        --------------------------------------SLXWTXXXXEXHTTTALCEGCS 22
MYDH048_        -------------------------------------------------------SAGVA 5
MYDH007_        ------------------------------------------------FXXLMLLPARRL 12
L16_            ----------------------------------------------PLAAQPSWVCRSSC 14
TLHC016_        ------------------------------------------------------------
TLHC022_        ------------------------------------------------------------
```

```
TLHC021F_       DLK-ISQTMPKTSSYNLTSLYA-LLVPDLTTIQGKVYCIGPG-QAFYATGDIIGDIRQAH 79
TLHC004_        ALK-ISQTMPKTYLYNLTSLYA-LLVPDLTTIQGKVYCIGPG-QAFYATGDIIGDIRQAH 79
MADH006_        DLK-ISQTMPKIYSYNLTSLCK-LLVSDLATIQEKVRRIGPG-QTFYATGDIIGDIR-AH 79
TLHC018F_       ICK--YHKQCQNHNSTIYCACKNSLYQDLTTIQGQVYCIGPG-QTFYATGDIIGDIRQAY 75
TLHC007_        ICK--YHKQCQNHNSTIYCACKNSLYQDLTTIQGQVYCIGPG-QTFYATGDIIGDIRQAY 75
TLHC037_        DLENISNNGAKIYLYNLPRLCT-LLVSDLITIQEKVYCIGPG-QTFYATDDVIGDIRQAY 82
TKMH002_        LGSANITNNAKNIIVQFTKPVKINCTRPNNNTR-KGIHIGPG-HAFYATDDIVGNIRDAY 69
MYDH036_        -----------------------------------GYDIGPRPNHFYAQEDIIGDIREAH 25
TLHC005_        SICKYHKQCQIYTSTTCSACENWIVPDLTTIQE-QVYCIGPG-QTFYAGG-IIGDIRQAY 74
MYDH004_        RICKYHKQCQNYNSTACQSCGNCVYQDLPTIQE-KVVRIGPG-QTFYATGDIIGNIRQAH 78
MYDH037_        RSRKYHKQSPKLYSYNLTSLVTISCIRPNNNTR-TSIPIGPG-QAFYRTGEIIGDIRQAH 79
TLHC016F_       SDLKISQTMPKISSYNLLEPVQINCTRPNNNTR-RSVHIGPG-QAFYATDDIIGDIRQAY 78
TLHC002_        LDLKISQTMPKIYLYNLLEPVQINCTRPNNNTR-RSVHIGPG-QAFYATDDIIGDIRQAY 78
MYSL032_        YNLKISQTMPKIYIVQLDQPVKITCTRTWQQYKNKVYRIGPG-QTFYATGDIIGDIRSAY 80
MYDH063_        RIWRISQTTPKLSSYNLRILCKLIVPDLTTIQD-KVYISGPG-QAFYGMGDIMGDLRQAI 83
MADH010_        RSENISANGTNIIVQLGNPVE-INCIRPNNNTRKSIPQSGQVKAILCNRLTLIRGYNDKL 81
MYDHX5_A1       --------------------------------------------------------NAKL 4
MYDH050_        RICNITNNAKTIIVQLVTPVK-INCTRPNNNTIKGV-HIGPGLSILCNR-SHNRGYKTSI 78
TLHC007F_       MICKYHRQCQKYNSTTSRVCTNSLVLRPSNNTRTSV-RIGPGQTFYATR-CCTGNIRQAY 78
TLHC017_        MICKYHRQCQKYNSTTYRVCTNLLVLRPSNNTRTSV-RIGPGQTFYATR-CCTGNIRQAY 78
TLHC012F_       RSANITDNVKNIIVQLTESVP-IHCFRPRNNTRTSV-RIGPGQTFYATG-AVTGNIRQAY 77
TLHC033_        DLENITDNAKNIIVQFNTSVG-INCTRPNNNTRKSI-TIGPG---------------KAL 59
MYDH046_        DLENITNNVKNIIVQLVNPVV-INCTRPNKNKRTRI-HIGPGRAFYTTS-AITGDIRKAY 79
MYDH039_        DLENITNNAKTIIVQLDKPVK-INCTRPNNNTRTSI-SIGPGQVFYRTG-DIVGNIRQAH 83
L12_            -IWNISDNAKTIIVQLAKPVS-INCTRPNTIQGKVY-IIGTRTAFYATG----------- 59
L25_            LDLKISQTMPKTSSYSLKRLYKLIVPDLTTIQE-KVYCIGPGMYSMQQVTYSEI------ 68
TKMH019X_       LDLKISQTMPKLYSYNLTRLCKLIVPDLTTIQE-QVYISDQDEHSMQQAASNRG------ 66
MADH014_        SIWKISQTMLKPSLYTLTRLSKLIVPDLTTIQG-EVYISDQDEHSMHRRHNGEISGSTLS 83
MYDH008_        SDLKISQTMPKIYSYNLTSLCKLIVPDLTTIQE-IWGLXXXGXXXYXXDXXGALRIAX-- 78
TLHC010F_       LDLKISQTMLKTYSYNLTHLWESIVPGPNNNTR-KSITIGTRESILCNR----------- 70
TLHC011F_       LDLKISQTMPKLSSYNLSLPVKINCSRPHNNTR-KSILHRTRTSILCNSVPVTGDIRKAH 79
TLHC024F_       SDLKISQIMPKLYSYNWLILLIINCTRPNNNTC-TSILFVDQDKHSMSS-DVIGDIRQAH 79
TLHC036_        SDLKISQIMPKLYLYNWLILLIINCTRPNNNTC-TSILFVDQDKHSMYS-DVIGDIRQAH 79
```

```
TKMH042F_           SICKYHKQCQNHN-STTCPACS-SLYQTSQQYK-KKYTVSDQDRHSMQQVTYS-EIQDKR 78
TLHC003_            SICKYHKQCQNHN-STTCPACSNYLYQTYQQYK-KKYTVSVQDRHSMQQVTYS-EIQDKR 79
TLHC011_            TICKYLKQCQNYN-STTCQTCANSLYQTSQQYK-KKCRYRTRTEHSMQQVTYS-GMSDKH 77
MADH005_            GSENITNNVKTII-VQLVNPVEIECTRTFQQYK-KKSYVSGQDKHSMQQVTSW-GISDRH 77
L26_                KICKHYKQCQKYN-STTSQACKNYLYQDLTTIQ-EQVYVSDQDKYTMQQGMSS-GISDKH 72
TLHC0024_           SDLRHLRQCSNHN-STISRDCNNSLYQASQQYK-EKYTYRDQDKHSMQQATSS-GIYDKH 78
TKMH017F_           SDLKISQTMAKISYVQLASLVNITCIMTYQSYK-KKYTVSDQGYHSIQQVISS-GIYDKH 72
TLHC014F_           RISKISQTMGQKYNSTTCLACEHYLCQTSSQYK-KKYTV--------------------- 62
MYDH027_            RSGKYLRQCSN-HNSTLYPDCKNYLYQTYQQYK-EKYTYRTRTNILCTGDIMG-KYKEAH 81
MYDH033_            RICKYHKQCQK-YNSTTCPACENYLYQTWQQYK-KKYAVSDQDKHSMQQVTSS-GISERI 80
L15_                SISKISQTIAKTIIVQLVKPVENSLYQTYQYYK-KKYNCSDQXFSFYATRCHNRGYKKSX 85
TLHC039X_           TICKVSQTIAKTIIVQLVKPVKINCTRPNNNTR-KSIRIGPG-QCILCNRCHNREYKYSI 79
TLHC018_            RSENITN-NAKTIIVQLVTACENSLSQTSQQYK-KKYTASDQDKHTLQLGARNRGYKKST 78
MYDH021             ------------------------------------YVWPGQ---PLYAQVHNWGYKKST 21
MADH004_            SHCKYSQTIDQTILVHSCRPET-SFGQPGPSNK-TKLTKSNSLSTVFKGTPGV------- 68
MADH003_            --CNPKSALQQLICTMTSQSHKAKVV----LYR---------DQYH------CLCYANSV 53
MYDH026X_           RCCCERTLLSRLYALPLVDNCDPRIASLPRLWRHSRCFVTSVDACHLTCILCCMCIIGES 82
MYDH040F_           ---LLC-------PPLRVPSSVKRPGSTAVLKYNRRILLCCSNYEVWCLEDLLLRSH-QI 70
MYDHX4_             ------------------------------------------NY---------------- 2
TLHC023_            ---LICSCVIWLHAAIDIPSAIKQ---TCCITIG-KILLHNSNYVLSFLN-LLLRSYSRL 74
MYDH053_            YHLLSYHLTPACCWAFQVLLGRSKCYSTGLMCYNRKILLYNCNYVLLFLY-PLLWSYQRL 82
MYDH019_            ------------LVVVTPSAIKAKCYSIGLLCYNRKILLYNSTYVLLFLD-PLLXFSSRL 71
MYDH025_            ------PSAICCIPSAIIPSAIKSSCCITIEKFSSTIKTMCCDFYIPSCG-----VSFYY 72
MYDH021_            -------------FSVVLVLLFQVLLFQLLLNRPDVLQSKNSPPQLKLCVVISRSPPEEL 75
MYSL033_            -------------CWHCNSSCYSKCYSKGLMYYN----TKNSLSQLKLCVVISRFFPEEL 68
MYDH042_            GSLANG--FNS---TWN-----ITANMQKLNG-----------TGPNDTITLPCRNKANY 54
MYDH028_            GDIRPSTIVNVSKLEWKKSLQKVARQLREYFNNKTIISCELLRKGIKKSQHIVLIVEENF 84
TLHC006_            SXFETG--------------XXGEXXXX---ERXXPIKKGPX-PVX--KDTXTXVARXRX 55
MYDH045_            ELFKACXSLTC----IQWXLXSGDXHGT---SXVCPSHXXEX-YLP--XDQXTXXARDIR 102
MYDH015_            SLHHYES----------MLELRHACASA---RTLATIGWGPS-CLP--DMLGIISAKFFC 68
MYDH050F_           XYCXSCPCGXXXXLTIXICILDQXXXLX---STLXTSXVAXLXSXEXLKNLXTXXSEGXF 117
MYDH038_            NSLLCITM--------SSISVLASLASA----RTLALWLGPPT-PWHAGIISSNVAPVPN 67
MYDH038_A1          NSLLCITM--------SSISVLASLASA----RTLALWLGPPT-PWHAGIISSNVAPVPN 67
MYDHX3_             ------------------------LAAA----SIIGLFI-----------------P--- 12
MYDH024_            XSLXXXXA--------XXXTXQQYMTHPX---RXLQLIX--IS-XXXQESISSQGTXNXX 66
MYDH052_            AXXEASXV--------DXMPLRGPPLXRXGPSXXXXXXXXSHXQRSSQGTFQSMCVTYLH 66
MYDH014_            DX--SCXX--------EXLPYQEX-XHX---SXLLPLMLVLX--KESQGTFQSMCVTYLH 55
MYDH049_            TT--RAAC--------EDLPYQEGALLF----DKDPDYSHCQ---GESRNFSKHVSHLPA 56
MYDH002_            GN--SCCC--------EVLPYQEGAGLV----DKEPLTIVTDX--GXXXELFKXCCLX-- 64
MYDH048_            TS--------------AAIPSATR-TIQ----QCMPLPS---------KDWTHPSCLLQC 37
MYDH007_            LG--------------ELQQVGGRGPEE----GCLPCGNSSG------QDWTHPSCLLQC 48
L16_                PVLMLRGP-----------PLSRRGPSLP---DSTPGLSSPN--LGKSQVTVKSMSKSLT 58
TLHC016_            ------------------------------------------------------------
TLHC022_            ------------------------------------------------------------

TLHC021F_           CDVNKTEWNNTLHKVAKQLRTHFTNKTIIFSSLLRRRYRNHNTSFYLW 127
TLHC004_            CDVNKTEWNNTLHKVAKQLRTHFTNKTIIFSYLLRRR------------YRNHNTSFYLW 127
MADH006_            IVRSINQNGITLYKRSLNNSKNTSGTNHNIYSLLREG------------ISKSHPSVYIV 127
TLHC018F_           CNVSRSAWNRTLQQVAKQLRDKSLRIKQ-------------------------------- 103
TLHC007_            CNVSRSAWNRTLQQVAKQLRDKSLRIKQSSLLNPQEG------------ISNLQHIALIV 123
TLHC037_            CNVSRSKWNDTLRQVAKQLGKYLGTTQELNLLTPQEG------------SRSPTHTFNCR 130
TKMH002_            CNVSREEWNKTLHQVAEQLRE-HFNNKTIVFNSSSGG-----------DLEITTHRFNCG 117
MYDH036_            CNSSRSDWNTTLIQVVKQLRR-YWN-TTIILNSSSGG-----------DLESTTHSCNCG 72
TLHC005_            CNVSKSEWNETLKGVVERLRDYFGNNATIIFANSSGG-----------DLEITTHSFNCG 123
MYDH004_            CNVSKANWTETLKLIAGQLETHFKN-KTIIFNSSSGG-----------DLEITTHSFNCG 126
MYDH037_            CNVSEAEWNETLQGVVDQLRKHFGN-KTIQFTNSSGG-----------DLEVTTHSFNCA 127
TLHC016F_           CIVNGSAWNKALQQVAGQLGK-IFVNKTIRFDKPSGG-----------DIEITTHSFNCR 126
TLHC002_            CIVNGSAWNKALQQVAGQLGK-IFVNKTIRFDKPSGG-----------DIEITTHSFNCR 126
MYSL032_            CQVNKSEWNNTLQKVARTIRR-ILWEHTYYFVNSSGG-----------GFRNHNPSFKWW 128
MYDH063_            VMS-LDQMGCNFRTSSWPLRN-TLGHKTLFFDNSSGVSK---------NPPLLLMGRIFF 132
MADH010_            I-------------VNCQLNQKM---------------------------DYN------- 94
MYDHX5_A1           ------------------------------------------------------------
MYDH050_            LSCQCSEWDNTLQQVAIQLRKYFGNRTRPIRIHYLLRRR---------LRNHNTSFVIVA 129
TLHC007F_           CNLNKTTWHATLPKGSQPIKN-ILYQNN--KLYQILRRG---------SRNNNTSFS--- 123
TLHC017_            CNLNKTTWHATLPKGSQPIKN-ILYQNN--KLYQILRRG---------SRNNNTLFS--- 123
TLHC012F_           CTINQTAWNATLQKVANQLRTYFVTQNN--KLYQILRRG---------SRNNNTSFS--- 123
TLHC033_            SICG-RESDTAMY------------------APP-------------------------- 74
MYDH046_            CVVN-RQTGITLYNRQLHNYENTL-TEQSYLKPPQEGIW---------KSHHIVLIVEEN 128
MYDH039_            CNVRCKQNGRRLYKGSLISYENTLGTKQYNLLTPQEGIS---------KSQHIALIVQES 134
L12_                CVIG--ALGQAF------------------------------------------------ 69
```

```
L25_            YDKH---------IVMS-VKQHGIKLC----------------------NRLPTDSEHT- 95
TKMH019X_       YKES---------ILCCQWIRMGITPS----------------------NRVATQ----- 90
MADH014_        YQSDQTGNNTFLQVVTPFKKILGTQPY----------------------SLIAPQEG--- 118
MYDH008_        XXGX--------CIXXGXTXXXGSXPCXSTILXDPSMQXCYLXRRGPRTSXMAXXXRRSX 130
TLHC010F_       ------------------------------------------------------------
TLHC011F_       CNIS--------RVDLEQHFICGKSTIR---------------------RTISEHNNNIY 110
TLHC024F_       CNIK--------WIKMESNFTTG---------------------------SLHQFNGSTC 104
TLHC036_        CNIK--------WIKMEYNFTTG---------------------------SLHQFYGSTC 104
TKMH042F_       IVTSVRQNGIELYNRSLHNSENTS-TKRS-FLRNPQEKI----------SKSQHIVLIVE 126
TLHC003_        IVPSVRQNGIELYNRSLHNYENTS-TKRY-FLRNPQEKI----------SKSQHIVFIVE 127
TLHC011_        IVMSVDLNGMKLYKGSLINYKNTSRTKQYSNLIAPQEGI----------SKSQHIVLIVE 127
MADH005_        IVMSVKQTGLKLSNCSLANSKHTLRTKQS-YLIAPQEGI----------SKSQHIVLIVE 126
L26_            LVRSINQNGIHLTTGRHPIKKNTFGEYNNNILINPQEG-----------IS--------- 112
TLHC0024_       IVMSVDQNGNKTLQQVAKKLRDTWTQHS--FLLTPQGGG----------LSNLQHIGFKL 126
TKMH017F_       IVMSANQNGMTLSGRSLNNSENTLGTTQSNLLTSSGGGS----------RS--HNTYFYL 120
TLHC014F_       ------------------------------------------------------------
MYDH027_        CNISRIRLGITLSYRSLHHLRRSLEHNHN--LSSLLRRG----------LEITTHWFKWG 129
MYDH033_        LSGQSIRMESHFTKGSSTIRRILREQTII--FTNSSGRG----------SRNHTHRFKLW 128
L15_            LQCQSIRMESDFTTGSLNNSADTLGTKQS-YLLTPQEWI----------YEITTHSLIXX 134
TLHC039X_       LSYQQNRLESHLSSGSLNNSENTLRTKQS-YLLTPQEGV----------SEVTTHSFNCG 128
TLHC018_        LYYQQSRLGITLYMRYVNDSENNLRTYQY-YLLTPQEGI----------SKSQHICLIVE 127
MYDH021_        L----------------------------------------------------------- 22
MADH004_        ------------------------------------------------------------
MADH003_        DH--------EVGANKYKLY--LL------------------------------------ 67
MYDH026X_       GYGTSTCVLVSMTRMVYNFYNRSLNSLRQHFQNITIA------------ISTSSGGDLEI 130
MYDH040F_       LLCSPR-RILIVELPPVELHYILX-IXSTTSCSIXX-------------HYXPLASKDGX 115
MYDHX4_         ------------------------------------------------------------
TLHC023_        LLCSS--IFLIVSLPVLKWYSSLI-YCYYNVLPLYL-------------LLCPPVHRNG- 117
MYDH053_        LFCFPQKRFSYLFATSWKSFIPLGPLMLQYAWPYIP-------------LLCTCGHKNA- 128
MYDH019_        LFYCSRRLSSLVCLPRSKVSFHSALLTGPICLLNIP-------------LLCHLLHRMLR 118
MYDH025_        MFCSPKVFSSLFSYLLSSFISILIQKTLPMCLVLISS------------YSLPPVGQRML 120
MYDH021_        VKIIVLFLKWCSPNCLATCCKVSFQLVYCHYNVLFLY------------PQLCTCASNGW 123
MYSL033_        ANIIVLFLKYP-LNCLATFCKSAILFCFIACHMRFYI------------SIWHLWIEGLV 115
MYDH042_        KYVAEPG--QAMYAPPS--KDLSLGGG--------------------------------- 77
MYDH028_        SNCNTSGLFNSTWDNNSTWGNISMQGLNSTDSITLPC-------------RIKQI----- 126
TLHC006_        XNGTXSKHSXXLFFX-------X-VAPG-------------------------------- 75
MYDH045_        IKDITXPXQXXXWRX-------XXXIXS-------------------------------- 123
MYDH015_        ASGSEKGDSTIL------------------------------------------------ 80
MYDH050F_       XPXXXXQNAXXXFXX-------SLFXYS-------------------------------- 138
MYDH038_        ALKIDLQSGFAFWTSNEPGIQLCHPVWLANSESFKEPVYV-------------------- 107
MYDH038_A1      ALKIDLQSGFAFWTSNEPGIQLCHPVWLANSESFKEPVYV-------------------- 107
MYDHX3_         ----------ACW---------C---WQ-------------------------------- 18
MYDH024_        FTEATNXTGXXLSSNX------C-PSHPRTXXXXVPP----------------------- 96
MYDH052_        SVGASGSG-GLAWNQQ------CMPLPSKDL----------------------------- 90
MYDH014_        SVVPGLWGIGMGTSNV------CPSHPR-------------------------------- 77
MYDH049_        FSGAWASG-GLAWNQQ------CMPLPSKG------------------------------ 79
MYDH002_        LXCHXXXX-XXXWG---------------------------------------------- 77
MYDH048_        MPLPSK------------------------------------------------------ 43
MYDH007_        MPLPSKXLXVXXXXSX------A------------------------------------- 65
L16_            FFRESLGVLVIGMEP-------AMDAPP-------------------------------- 79
TLHC016_        ------------------------------------------------------------
TLHC022_        ------------------------------------------------------------

TLHC021F_       RRIFLLCSSSLFNSTWM---------------------------------------- 144
TLHC004_        RRIFLLCSSSLFNSTWM---------------------------------------- 144
MADH006_        EENLLLESSGLFNSTWNPGSMPAC--------------------------------- 151
TLHC018F_       ---------------------------------------------------------
TLHC007_        EENFSIVIHQPVSSHLGKSLAFGKFTVCGKCHFTSPW-------------------- 160
TLHC037_        GEFFY---------------------------------------------------- 135
TKMH002_        GEFFYCTHQGYFIAGWA--------DNASMQTIARGKCPLLSMRISNINMWRDRQP- 165
MYDH036_        GEFFLL--------------------------------------------------- 78
TLHC005_        GEFFYCNTSNL---------------------------------------------- 134
MYDH004_        GEFFYWNHQGLFNSTWTLPAYSKF--------------------------------- 150
MYDH037_        GEFFYCNTSGLFKS-FTMPAAGVITDANTLYSPCGSTILICAESRQYDP-------- 175
TLHC016F_       GEFFYCNHQACLIALGL---------------------------------------- 143
TLHC002_        GEFFYCNHQACLIALGL---------------------------------------- 143
MYSL032_        GES-SMESSGLFNSLES---------------------------------------- 144
MYDH063_        GSWLFNVAFRARGKNGG---------------------------------------- 149
MADH010_        ---LLHKGSPETGSRQSL--------------------------------------- 109
MYDHX5_A1       ---------------------------------------------------------
```

```
MYDH050_            GEYLLLNTSDLSNSLGILAL------------------------------------- 149
TLHC007F_           ---LCRRGFSIVTHQACLIALEWHCSMQELIHGRYTLHAESSKFS------------ 165
TLHC017_            ---LCRRGFSIVTHQACLIALEWHCSMQELIHGRYTLHAEYSKFY------------ 165
TLHC012F_           ---LWRES------------------------------------------------- 128
TLHC033_            ---------------------------------------------------------
MYDH046_            FLLGPSGCSYLGSYYTAAG---VNAHFTLHPIGNFRW-------------------- 162
MYDH039_            FSIVIHQVCLIVYNASGRSNYMTQIAFTLHANKQYNMCKSGKHDP------------ 179
L12_                ---------------------------------------------------------
L25_                ------LRTQQS--------------------------------------------- 101
TKMH019X_           ------LREHF---------------------------------------------- 95
MADH014_            ------FRNYHPLVSMG---------------------------------------- 129
MYDH008_            MEAXACXRXXIXXXHAXXGXMPAXTSFXXGNXGGXRRCP------------------ 169
TLHC010F_           ---------------------------------------------------------
TLHC011F_           SLLRRGSRNHNTYVSLWRRSFFYCNTSGAWLIALESCSHRESHCQHAGGQMARGQN- 166
TLHC024F_           ---------------------------------------------------------
TLHC036_            ---------------------------------------------------------
TKMH042F_           ESFSIVIHQACLIAL------------------------------------------ 141
TLHC003_            ESFSIVIHQACLIAL------------------------------------------ 142
TLHC011_            ENFSIVIHQAWLTAPGPITPRSQWDGARLKFTSMQ---------------------- 162
MADH005_            ENSFIGIIKACLIALGHYQHTANS--------------------------------- 150
L26_                ---EFTTH------------------------------------------------- 117
TLHC0024_           WEGEFSYWNT----------------------------------------------- 136
TKMH017F_           WRRIFLLYSSGPVSQHLGESHLQSKWQHGGEANARYSLCGSAMEYV----------- 166
TLHC014F_           ---------------------------------------------------------
MYDH027_            ---------------------------------------------------------
MYDH033_            RRIFYWNHQACLIALGIPVPCQL---------------------------------- 151
L15_                -RIXLLNHQDWLMALXYXTAXXXQIXRANEXXPPXDKQLX----------------- 173
TLHC039X_           GRVFLL---------EYIRTCLSSILYHND--------------------------- 149
TLHC018_            EKFFLLSYIRSLVNSPGILIAPGITLTTCRGSNGTGSK------------------- 165
MYDH021             ---------------------------------------------------------
MADH004_            ---------------------------------------------------------
MADH003_            ---------------------------------------------------------
MYDH026X_           TTHSLIVEENFSIVTIRPVNSPWNAMQPCGIN------------------------- 162
MYDH040F_           AGXXXYXSIVGRTGXTNR--------------------------------------- 133
MYDHX4_             ---------------------------------------------------------
TLHC023_            SAWVLYR-IL----------------------------------------------- 126
MYDH053_            WPLVLY--------------------------------------------------- 134
MYDH019_            PGHKGHLL------------------------------------------------- 126
MYDH025_            WPWVL---------------------------------------------------- 125
MYDH021_            PGQT----------------------------------------------------- 127
MYSL033_            LGLFRH--------------------------------------------------- 121
```