5-2013

EVOLUTION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 CLADE C ENVELOPE V1-V5 REGION DURING DISEASE PROGRESSION IN NON-HUMAN PRIMATE MODEL

For Yue Tso
University of Nebraska-Lincoln, s-ftso1@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/bioscidiss

Part of the Immune System Diseases Commons, Virology Commons, and the Virus Diseases Commons

Tso, For Yue, "EVOLUTION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 CLADE C ENVELOPE V1-V5 REGION DURING DISEASE PROGRESSION IN NON-HUMAN PRIMATE MODEL" (2013). Dissertations and Theses in Biological Sciences. 48.
http://digitalcommons.unl.edu/bioscidiss/48

This Article is brought to you for free and open access by the Biological Sciences, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Dissertations and Theses in Biological Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
EVOLUTION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 CLADE C ENVELOPE V1-V5 REGION DURING DISEASE PROGRESSION IN NON-HUMAN PRIMATE MODEL

By

For Yue Tso

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Biological Sciences

Under the Supervision of Professor Charles Wood

Lincoln, Nebraska

May, 2013
Human immunodeficiency virus type 1 (HIV-1) clade C strain is the fastest spreading HIV-1 strain globally, especially in Africa. It has been decades since the acquired immune deficiency syndrome (AIDS) pandemic first started. However, an effective anti-HIV-1 vaccine is not yet available, which is partly due to the highly variable nature of HIV-1 envelope gene and the absence of a suitable animal model. Strengthening of the understanding of envelope evolution during disease progression will contribute significantly towards future anti-HIV-1 treatment and preventions.

Non-human primates have been an essential animal model for many biomedical research areas. Using simian-human immunodeficiency virus (SHIV) expressing the HIV-1 clade C V1-V5 envelope regions from Zambia and infecting rhesus macaque, we demonstrated close genetic evolution of the HIV-1 envelope in both non-human primates and patient during disease progression. Utilizing this animal system, we further examined the evolution of the biological properties of HIV-1 envelope and its association with disease progression. In addition, we applied next generation sequencing technology to analyze the genetic evolution of HIV-1 envelope in a different non-human primate, pigtailed macaque, which was infected with a closely related SHIV strain. Our analysis
showed that HIV-1 envelope underwent similar genetic evolution through the disease course in three different primate species; the human host and two non-human primate species.

In summary, our studies have validated the relevance of SHIV infected non-human primates as the ideal animal model for HIV-1 research. In addition, we have also chronicled the changes in biological characteristics of the HIV-1 envelope through disease progression. Thus, our results underscore the dynamic evolutionary relationships between the intrinsic properties of the HIV-1 envelopes, their evolution and host immune response. Findings from this study may contribute to the eventual resolution of the AIDS pandemic and our knowledge of the roles that HIV-1 envelope plays during disease progression.
Acknowledgements

I would like to thank my mentor, Dr. Charles Wood, for the opportunity to be a graduate student in his laboratory. His tireless patience and encouragement is instrumental to the completion of this dissertation. Over the years, he has offered selfless help and invaluable guidance to not only me, but to other graduate students under his supervision. I have learned tremendously through his enthusiasm in sciences and thought provoking discussions, which will be an indelible part of my training as a scientist.

I would also like to thank my committee members Dr. Jack Morris, Dr. Fernando Osorio, Dr. Peter Angeletti and Dr. Clinton Jones for their thoughtful discussions and guidance throughout the course of my study.

The completion of this dissertation will not be possible without the supports from many past and present members of Wood Lab. In particular, I deeply appreciate the friendships of Danielle Shea, Kay Crabtree, Pam Snyder and Eun Hee Kwon; Drs. Sandra Gonzalez-Ramirez, Damien Tully, Veenu Minhas, Pankaj kumar and Dawn Eggert. Their helping hand is always assured in times of adversity. The priceless memories of fun road trips and events that we shared will always brighten my day. In addition, I want to specifically thank Drs. Levon Abrahamyan and Zhang Hong for their guidance and unflagging support in my research.

I dedicate this dissertation to my parents and sister, who have unwaveringly supported me through the years. They have provided a loving and stimulating learning environment that allowed me to accomplish this achievement. For this, I am eternally grateful. I want to express my utmost gratitude and appreciation to them.
# Table of Contents

Abstract ..........................................................................................................................................................i

Acknowledgments .........................................................................................................................................iii

Table of contents .........................................................................................................................................iv

List of abbreviations ...................................................................................................................................viii

Chapter 1: Introduction and literature review ....................................................................................1

History and epidemiology of HIV-1 .......................................................................................................2

HIV-1 genome and structure ....................................................................................................................5

HIV-1 proteins and functions ....................................................................................................................7
  Gag (MA, CA, NC, p2, p1 and p6) .........................................................................................................8
  Pol (PR, RT and IN) ..............................................................................................................................9
  Accessory proteins (Vif, Vpr/Vpx, Vpu, Nef) ....................................................................................11
  Tat and Rev ...........................................................................................................................................15
  Envelope ................................................................................................................................................16

Pathogenesis of HIV-1 ..............................................................................................................................19

Life cycle of HIV-1 ....................................................................................................................................19

Transmission ...............................................................................................................................................21
Infection………………………………………………………………………….23

Disease progression…………………………………………………………………25

Non-human primate models………………………………………………………..28

SIV and SHIV models……………………………………………………………..28

stHIV-1 model……………………………………………………………………..31

Research aims……………………………………………………………………….33

Figures and Figure legends…………………………………………………………36

Chapter 2: A comparative study of HIV-1 clade C env evolution in a Zambian infant with an infected rhesus macaque during disease progression……………….40

Abstract………………………………………………………………………………41

Introduction…………………………………………………………………………43

Materials and methods……………………………………………………………45

Results……………………………………………………………………………….48

Discussion………………………………………………………………………….53

Acknowledgments………………………………………………………………58

Figures and Figure legends…………………………………………………………59

Chapter 3: Variations in the biological functions of HIV-1 clade C envelope in a SHIV-infected rhesus macaque during disease progression…………………………68
Chapter 4: Dynamics of envelope evolution in clade C SHIV-infected pig-tailed macaques during disease progression analyzed by ultra-deep pyrosequencing

Abstract..............................................................................................................99

Introduction.........................................................................................................100

Materials and methods......................................................................................103

Results................................................................................................................107

Discussion..........................................................................................................115

Acknowledgments.............................................................................................120

Figures and Figure legends..............................................................................121

Supporting information....................................................................................141

Chapter 5: Concluding remarks........................................................................144
List of Abbreviations

AIDS: Acquired immune deficiency syndrome

APOBEC3A: Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3A

APOBEC3F: Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3F

APOBEC3G: Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G

AZT: Azidothymidine

BST-2: B cell stromal factor 2

βTrCP: Beta-transducine repeat containing protein

CCR5: Cysteine-cysteine chemokine receptor 5

CD: Cluster of differentiation

CDC: Centers for disease control and prevention

CPM: Counts per minute

CTL: Cytotoxic T-lymphocytes

CXCR4: Cysteine-X-Cysteine chemokine receptor 4

dN: Nonsynonymous

dS: Synonymous substitution

Env: Envelope glycoprotein
ELISA: Enzyme linked immunosorbent assay

ER: Endoplasmic reticulum

FDA: Food and drug administration

FRET: Fluorescent resonance energy transfer

Gag: Group-specific antigens

GALT: Gut-associated lymphoid tissue

HAART: Highly active anti-retroviral treatment

HPCs: Hematopoietic progenitor cells

HIV-1: Human immunodeficiency virus type 1

HIV-2: Human immunodeficiency virus type 2

HLA: Human leukocyte antigen

Hrs: Hours

HTLV: Human T-lymphotropic viruses

I-domain: Interaction domain

IN: Integrase

KS: Kaposi’s Sarcoma

KSHV: Kaposi’s sarcoma-associated herpesvirus
MA: Matrix protein

mAbs: Monoclonal antibodies

M-domain: Membrane binding-domain

MHC-I: Major histocompatibility complex class I

MHC-II: Major histocompatibility complex class II

MLA: Monkey leukocyte antigen

MLV: Murine leukemia virus

Nef: Negative factor

NF-kB: Nuclear factor kappa beta

NK: Natural-killer cells

NNRTI: Non-nucleoside reverse transcriptase inhibitor

NRTI: Nucleoside reverse transcriptase inhibitor

LAV: Lymphadenopathy-associated virus

LDPCR: Limiting dilution PCR

LTR: Long terminal repeats

PBS: Primer binding site

PBMC: Peripheral blood mononuclear cell
PCR: Polymerase chain reaction

PCP: Pneumocystis carinii pneumonia

PNGS: Potential N-glycosylation sites

PIC: Pre-integration complex

PR: Protease

Pol: Polymerase

PPT: Polypurine tract

Pt: Post transfection

Pi: Post infection

Rev: Regulator of expression of viral genes

RRE: Rev response element

RT: Reverse transcriptase

SAMHD1: SAM- and HD-domain-containing protein

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SIV: Simian immunodeficiency virus

SIVcpz: SIV of chimpanzee

SIVgor: SIV of gorilla
SIVsmm: SIV of sooty mangabey

SHIV: Simian-human immunodeficiency virus

Sp1: Specificity protein 1

stHIV-1: Simian-tropic human immunodeficiency virus type 1

SU: Surface subunit of HIV-1 envelope

TAR: Trans-activation response element

Tat: Transcriptional transactivator

TFBS: Transcription factors binding sites

TM: Transmembrane domain of HIV-1 envelope

tRNALys: Transfer RNA lysine

TRIM: Tripartite motif

UDPS: 454 ultra-deep pyrosequencing

UNAIDS: Joint United Nations Programme on HIV/AIDS

vDNA: Viral DNA

Vif: Viral infectivity factor

VLP: Virus like particles

vRNA: Viral RNA
Vpr: Viral protein R

Vpu: Viral protein U

Vpx: Viral protein X
Chapter 1

Introduction

A better understanding of the human immunodeficiency virus type 1 (HIV-1) envelope evolution during disease progression is critical to the development of effective anti-HIV-1 vaccines and treatments. However, the lack of suitable animal models and variation in methodologies has been the major hindrance to achieving this goal. The aim of this dissertation is to address these issues by focusing on the evolution of HIV-1 clade C V1-V5 envelope regions during disease progression. We studied the simian-human immunodeficiency virus (SHIV) infection of non-human primates as the animal model, in order to focus on the genetic and biological functions variations of the HIV-1 envelope through the development of acquired immune deficiency syndrome (AIDS).

This dissertation is separated into five chapters. Chapter one includes the literature review that is categorized into five sections. These sections will provide fundamental information ranging from the history and epidemiology of HIV-1 to the various non-human primate models currently available for HIV-1 research. Through these reviews, the readers will be able to better comprehend the concept of our studies. From chapter two to four, the research data and results of this dissertation will be presented as published in their respective journals. In chapter two, we will show the similarities in the genetic evolution of HIV-1 envelope in an SHIV infected rhesus macaque with HIV-1 infected patient using bioinformatics analysis. Thereby, it will demonstrate the relevance of our animal model for HIV-1 research. In chapter three, we will examine and chronicle the changes in the biological characteristics of HIV-1
envelope during disease progression by using the same animal model as in chapter two. We will establish the association of various biological functions of HIV-1 envelope within the context of disease progression. In chapter four, we further substantiate our previous findings by analyzing the genetic evolution of HIV-1 envelope in a different non-human primate species (pigtailed macaque) infected with a closely related SHIV. This dissertation will be completed with the concluding remarks of our studies in chapter five.

**Literature review**

**History and epidemiology of HIV-1**

In 1981, there was a sudden increase in reports of rare opportunistic infections such as Pneumocystis carinii pneumonia (PCP) and the AIDS-associated malignancy Kaposi’s sarcoma (KS) among young homosexual patients in California and New York (4, 179). Within the next few years, similar cases were also increasingly observed among other groups such as the intravenous drug users, heterosexual women and possibly the first recorded case of vertical transmission of HIV-1 from mother to child (1, 5, 252). These observations were not unique to the United States, but were reported around the world at a similar time (101, 319). Although the patients came from diverse backgrounds, they were all suffering from immune deficiency symptoms that seemed to be acquired rather than congenital. Hence, it gave rise to the term Acquired Immune Deficiency Syndrome (AIDS). Based on the information gathered from various cases, the scientific community at that time began to suspect that an unknown agent was behind the AIDS epidemic. More importantly, they realized that this unidentified agent is infectious and
blood or sexual contact with an AIDS patient was recognized as possible routes of transmission.

The first major milestone in HIV-1 research occurred in 1983. Dr Luc Montagnier and colleagues at the Institute Pasteur in France successfully isolated a new human retrovirus, lymphadenopathy-associated virus (LAV), from a homosexual patient with multiple lymphadenopathies (23, 263). Shortly thereafter, a new type of human T-lymphotropic viruses (HTLV), HTLV-IIIB, was discovered by Dr Robert C. Gallo and his colleagues using the samples obtained from Dr Montagnier’s group (120, 122). It was later determined that both LAV and HTLV-IIIB are in fact HIV-1. More importantly, Dr Gallo’s group first proposed and demonstrated HIV-1 as the causative agent for AIDS (121, 300, 321, 331, 341). These discoveries set the stage for the first commercially available diagnostic blood test for HIV-1 in 1984, which was credited for preventing HIV-1 transmission through blood donations. This was closely followed by other breakthroughs such as the cloning and sequencing of the full length HIV-1 genome (306, 402).

Towards the end of 1980s, azidothymidine (AZT) clearly demonstrated its effectiveness to increase the survival rate of AIDS patients in a double-blind clinical trial, and became the first Food and Drug Administration (FDA) approved antiretroviral drug available (110). Since then, many new drugs targeting various parts of the HIV-1 life cycle have been developed. Nonetheless, the severe toxic side effects, high cost of these drugs and the development of drug resistant viruses necessitate the development of anti-HIV-1 vaccine. However, an efficacious anti-HIV-1 vaccine has proven to be more difficult to achieve despite decades of research. Concurrently, human immunodeficiency
virus type 2 (HIV-2) and simian immunodeficiency virus (SIV) were discovered in African AIDS patients and macaque, respectively (75, 82). All these milestones established the foundation for the current HIV-1 research in general and this dissertation in particular.

HIV-1 is divided into several subgroups (M, N, O and P) based on phylogenetic analysis of the complete viral genome, with group M representing the major pandemic strain. The origin of HIV-1 remained uncertain for years until recent studies showed that HIV-1 were most likely derived from SIV of chimpanzees (SIVcpz) and possibly gorillas (SIVgor) (126, 393). However, the source of these SIVs is still controversial. Sequence analysis of SIVcpz revealed that it is comprised of several other SIVs (19). This suggests that chimpanzees might have come in contact and been infected with multiple SIVs in the past, potentially during the hunting of monkeys (19, 372). Similar to HIV-1, HIV-2 was reported to have originated from SIV found in sooty mangabeys (SIVsmm) (223). However, due to its relatively low pathogenicity, HIV-2 plays a minor role in this AIDS pandemic and infections are confined mainly to West Africa (88, 318).

The question of how this cross-species transmission of SIV into human occurs has fueled numerous theories. However, it is widely accepted that traditional hunting of non-human primates in Africa may have initiated the first zoonotic infection of SIVcpz in humans. Subsequently, viral adaption and evolution over time in the human host resulted in the birth of modern HIV-1. The urbanization of Africa and rapid globalization over the past 100 years may have propelled the spread of HIV-1 around the world and started the modern AIDS pandemic (68). By 2011, there were more than 33 million people living with HIV-1 worldwide according to UNAIDS (3). Sadly, 330 000 children were newly
infected with HIV-1 in 2011 alone. Nearly 70% of the HIV-1 patients lived in sub-Saharan Africa regions. Some countries such as Botswana, Lesotho and Swaziland have a HIV-1 prevalence rate of ~25% among their adults (3). This high HIV-1 prevalence rate created not only economic issues for these countries, but also a tremendous impact on their social structure with ~15 million orphans due to parents being lost to AIDS (2). Through monumental efforts invested in education, behavioral changes, availability of condoms and antiretroviral drugs, the number of AIDS-related deaths each year has been kept stable at around 2 million (3). Critically, increased accessibility of antiretroviral drugs for pregnant women have helped to decrease mother to child transmissions (3). Despite all these efforts, there are still over 2.5 million new HIV-1 infection cases each year (3). A practical and efficacious anti-HIV-1 vaccine is sorely needed to control this AIDS pandemic.

**HIV-1 genome and structure**

HIV-1 is a lentivirus member of the retroviridae family. It is a near spherical enveloped virus packaged with two single-stranded positive sense viral RNA (vRNA). The HIV-1 genome is ~9.5 kb nucleotides in length, with overlapping reading frames encoding multiple viral proteins (Fig 1A). The majority of the HIV-1 genome encodes its structural proteins group-specific antigen (Gag), envelope (Env), as well as polymerase (Pol) which encompasses three critical viral enzymes; protease (PR), reverse transcriptase (RT) and integrase (IN). Some of these proteins are expressed as precursor polypeptides before processing by PR into mature proteins. The viral genome also encodes several accessory proteins such as the viral infectivity factor (Vif), viral protein R (Vpr), viral protein U (Vpu) and negative factor (Nef) (Fig 1A). Although the accessory proteins are
considered non-essential for \textit{in-vitro} cultures, they play an important role for \textit{in-vivo} pathogenesis. The 5’ and 3’ ends of viral genome are flanked by identical non-coding long terminal repeats (LTR) (Fig 1A). The LTR contains regulatory elements such as transcription factors binding sites (TFBS), TATAA box and trans-activation response element (TAR) that are crucial for viral replication and transcription. Mutations within the TFBS can affect the viral replicative fitness in different cell types (55, 290, 395). Also, interaction of the viral transcriptional transactivator (Tat) with the secondary RNA structure formed by TAR is essential for replication and enhancing viral gene transcriptions (14, 346).

HIV-1 virions are approximately \~145 nm in diameter with an outer membrane consist of the host cell phospholipid bilayer derived during viral budding (39). The virion surface is embedded with the viral envelope (Env) spikes. Each spike is composed with homotrimers of the envelope surface subunit gp120 (SU) and transmembrane domain gp41 (TM) linked by non-covalent bonds (132, 246). The SU is exposed on the virion surface, while TM spans the phospholipid bilayer to form a stalk that connects the SU to the virion. Studies have shown that there are between 7 to 14 envelope spikes on the virion surface (66, 444). However, data from a portion of this dissertation suggests that the level of virus-associated SU content changes depending on disease progression and might exceed the maximum level proposed above. In addition to the viral envelope, many host cell surface proteins are also incorporated onto the virion surface during viral budding (67, 281).

Beneath the phospholipid bilayer lays the core shell, composed of the matrix protein (MA) (432). Located at the center of this core shell is the cone shaped
nucleocapsid formed by the capsid protein (CA). Importantly, HIV-1 virion buds from the host cell with an immature spherical core, only after maturation does it develop a cone shaped nucleocapsid (87). Assembled within this nucleocapsid are the two single-stranded positive sense vRNA that are tightly bound by the nucleocapsid protein (NC). Near the 5’ end of these vRNA are transfer RNA lysine (tRNA<sub>lys</sub>) molecules, derived from the host cell, acting as the primer for reverse transcription (306). Several viral proteins such as Vpr, reverse transcriptase (RT) and integrase (IN) are also located in the nucleocapsid. Moreover, a number of studies have demonstrated that host cellular proteins can also be packaged into the virion during the viral assembly process (67, 406).

Similar to HIV-1, SIV is also a lentivirus member of the retroviridae family. SIV shares similar genomic and virion structures as HIV-1 with a few exceptions (Fig 1B). One of these differences lies within the LTR. SIV contains an average of one nuclear factor kappa beta (NF-kB) and four specificity protein 1 (Sp1) TFBS, while HIV-1 has on average two NF-kB and three Sp1 TFBS. Variations in the number of TFBS were reported to affect viral pathogenesis (94, 107). More importantly, SIV encodes the viral protein X (Vpx) instead of Vpu that is unique to HIV-1 (Fig 1B). While HIV-1 and SIV have evolved naturally, SHIV is a laboratory-generated chimeric virus between SIV and HIV-1. To generate SHIV, the HIV-1 genes tat, regulator of expression of viral genes (rev), vpu and env are inserted into the pathogenic molecular clone, such as SIVmac239, to replace their counterparts (Fig 1C). The resulted SHIV virions are fully functional and infectious in non-human primate animal models (240).

**HIV-1 proteins and functions**
**Gag (MA, CA, NC, p2, p1 and p6)**

The *gag* gene (Pr55\textsuperscript{Gag}) is expressed as a ~55kDa polypeptide precursor that consists of the viral proteins MA (p17), CA (p24), p2, NC (p7), p1 and p6. Its main function is assembly of the virions. Gag alone can form virus-like particles (VLP) in the absence of any other viral proteins (91). During maturation shortly after viral budding, Gag undergoes proteolytic cleavage by PR into individual proteins. This step was reported to occur in a specific order and at different rates for each protein (191, 291). Variations in these cleavage events could have detrimental consequences on virus infectivity (78, 270). The virion core shell is formed by the MA, which serves multiple functions. After translation, MA is myristylated by the host cell enzyme N-myristyl transferase and directs the gag polypeptide towards the plasma membrane via its membrane binding-domain (M-domain) (41, 287, 340). Several studies have also demonstrated that MA interacts with gp41 of the viral envelope to facilitate Env incorporation into the virion (168, 426). However, it has no effect on the incorporation of Env with short cytoplasmic tails (113). Additionally, MA contains nuclear localization signals, suggesting that it may be involved in the import of the viral pre-integration complex (PIC) into the nucleus and export of viral RNA out of the nucleus (43, 103).

The cleavage of the MA-p2 polypeptide by PR releases the CA, which is the main component of the nucleocapsid. CA mainly functions in Gag-Gag interaction as well as incorporation of the cellular protein cyclophilin A, which is necessary for viral replication (35, 124, 125). Similar to CA, NC also participates in Gag-Gag interaction and virus assembly through its interaction domain (I-domain) (287). NC is a hydrophilic protein that is indispensable for packaging of vRNA into the virion. It binds to the
packaging sequence (Ψ site) within the vRNA, located between the 5’ LTR and gag start codon, via its zinc-finger domain (193, 325, 342). NC also promotes the binding of tRNA\(^{Lys}\) to the primer binding site (PBS) of the vRNA and enhances the reverse transcription process (229).

Lastly, the budding of virion from the host cell occurred through p6 via the endosomal sorting complex required for transport (ESCRT) pathway (171, 390, 413). In addition, p6 is required for the incorporation of Vpr into the virion (202). Unlike other proteins within the Gag polypeptide, the function of the smaller peptides p2 and p1 are not well understood, but may be involved in the regulation of Gag polypeptide cleavage by PR (409).

Pol (PR, RT and IN)

The pol gene encodes three pivotal viral enzymes as described earlier. The Pol proteins are first translated as Gag-Pol precursor polypeptides via a reading frame shift at the C-terminus of gag (184). It was suggested that this Gag-Pol polypeptide dimerized to form homo-dimeric PR, which was then released by autolytic cleavage (90, 415). The function of PR is solely to cleave the Gag and Gag-Pol polypeptides into their corresponding individual viral proteins. This cleavage step is critical for maturation and functionality of the virion. PR can recognize several cleavage sites and its specificity can be affected by the primary sequence as well as structure of the substrate (292). The detailed structure of PR has been resolved and describe elsewhere (273, 352).

One of the viral proteins released after cleavage of Gag-Pol polypeptide is RT, which is a hallmark of retroviruses. RT functions as the RNA-dependent DNA
polymerase and RNAse H. Its RNA-dependent DNA polymerase activity is responsible for the transcription of vRNA into double-stranded linear viral DNA (vDNA) upon viral entry into the cytoplasm (21, 330, 374). Whereas, it’s RNAse H activity degrades the vRNA template immediately after transcription (336). This entire process of copying vRNA into vDNA is known as reverse transcription (77). Briefly, RT utilizes the tRNA\textsuperscript{lys} bound to the PBS of vRNA as a primer to reverse transcript the viral 5’ LTR into a short minus-strand vDNA. Simultaneously, the RNAse activity of RT degrades the transcribed vRNA. Together with the RT complex, this short minus-strand vDNA was transferred to the 3’ LTR of vRNA for continuation of the reverse transcription. Almost the entire vRNA is degraded after the full minus-strand vDNA is synthesized, except for the polypurine tract (PPT) within the 3’ LTR that are resistant to RNAse H activities. RT utilizes the PPT as a primer for synthesis of the second plus-strand vDNA to complete the reverse transcription process. RT was first produced as a homodimer with two p66 subunits. However, PR cleaved one of these subunits into p51, generating a heterodimer RT with p66 and p51 subunits. The enzymatic activities of RT are possessed by p66, while p51 is believed to have a structural role only (167). Detailed RT structure has been resolved and described (185). Importantly, the lack of proofreading ability in RT resulted in its low fidelity. It has been estimated that 1 in 1700 to 4000 nucleotide was misincorporated into the vDNA during reverse transcription (301, 314). Surprisingly, this high error rate actually benefits the virus by increasing its genetic diversity, contributing to the viral escape from the host immune surveillance (227).

The pol gene also encodes IN which functions to insert the double stranded linear vDNA into the host genome (139). This was accomplished by the DNA cleavage and
strand-transfect abilities of IN in the PIC (139). The detailed steps of integration have been reviewed (76). Briefly, IN cleaves two bases from the 3’ terminus of vDNA that serve as the attachment site for host genome. After entry into the cell nucleus, IN facilitates the strand transfer of cleaved vDNA into the target host chromosome. Although the integration sites within the host chromosome are believed to be randomly selected, some studies suggest that it might occur at a higher frequency in regions with open chromatin structure (79). The structure of IN has been resolved and it exists as a homodimer with Z-finger domains (46, 105).

Due to their important roles in the virus life cycle, PR, RT and IN are targets for many antiretroviral drugs. For example, ritonavir inhibits the PR, non-nucleoside reverse transcriptase inhibitor (NNRTI) and nucleoside reverse transcriptase inhibitor (NRTI) inhibits the RT activities. A new class of drug, raltegravir, was recently developed to inhibit IN. Different combinations of these drugs constitute the highly active anti-retroviral therapy (HAART) for AIDS patients. Unfortunately, drug resistant mutants are rapidly rising in treated individuals, creating a major issue for anti-HIV-1 therapies (360).

**Accessory proteins (Vif, Vpr/Vpx, Vpu, Nef)**

HIV-1 encodes several accessory proteins, Vif, Vpu, Vpr and Nef. These proteins were once thought to be non-essential for *in-vitro* viral growth. However, recent studies have shown that they are critical for *in-vivo* viral replication and disease pathogenesis. Vif is a small ~23 kDa protein that accumulates within the host cytoplasm after translation (141). Several studies have demonstrated that Vif can also be incorporated, resulting in an estimated 60 to 100 molecules in the virion during viral assembly (235). It
was established that Vif is necessary for viral replication in primary T-cell but not in cell lines such as SupT1 (119). Later, it was discovered that the cellular protein, apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G (APOBEC3G) can inhibit HIV-1 replication in the absence of Vif (351). APOBEC3G is a cytidine deaminase with an unclear normal cellular function, but was reported to have a regulatory role in microRNAs (170). Alternatively, APOBEC3G could have evolved as a specialized cellular anti-viral defense. APOBEC3G restricts the HIV-1 replication by deamination of cytidine to uridine in the minus-strand vDNA, causing hypermutations within the viral genome. It is the function of Vif to effectively antagonize this negative effect of APOBEC3G, by binding and inducing the degradation of APOBEC3G via the ubiquitin-proteasome pathway (248, 249, 257, 351). Importantly, the inhibition of APOBEC3G by Vif is highly species dependent, such that APOBEC3G from non-human primate can only be inhibited by SIV Vif and vice versa for HIV-1 (33). Other members of the APOBEC family, such as APOBEC3F, may have similar functions as APOBEC3G and can also be inhibited by Vif (234).

Vpr is a ~15 kDa protein that is expressed by all HIV-1 and SIV. However, a closely related protein, Vpx is only expressed in HIV-2 and certain SIV lineages (Fig 1) (117). Vpr/Vpx bears multiple functions and is important for in-vivo pathogenesis. It was shown that Vpx is required for disease progression and deletion mutants of Vpr/Vpx displayed significantly delayed onset of AIDS (135, 160). One of the functions of Vpr is localization of the viral PIC into the nucleus of non-dividing cells such as macrophages (299). This ability vastly expands the number of permissive cell types for the virus to infect. Vpr was also shown to induce G2 cell cycle arrest by an unknown mechanism
Another function of Vpr is inducing T-cell apoptosis, resulting in by-stander T-cell death that might contribute to the decrease of CD4\(^+\) T-cell counts in AIDS patients (37). Moreover, a study had linked Vpr to the reduction of antigen presentation by affecting the infected dendritic cells maturation and cytokine production (242). Although Vpx shares similar functions as Vpr, there are some differences between the two proteins. For instance, Vpx cannot induce G2 cell cycle arrest but virus expressing Vpx can establish infection of macrophages by overcoming a restriction in reverse transcription (349). A study had demonstrated that Vpx can inhibit the anti-viral effect of restriction factor, APOBEC3A, in myeloid cells (26). Importantly, a recently discovered dendritic and myeloid cell specific HIV-1 restriction factor, SAMHD1, can be antagonized by Vpx (217, 218). Given the important role that dendritic cells have as antigen presenting cells, overcoming the SAMHD1 restriction allows productive infection by virus expressing Vpx and may increase the viral pathogenesis (48).

Vpu is a 16 kDa protein unique to HIV-1 and SIVcpz (Fig 1) (210). Vpu is an oligomeric type-1 integral membrane protein that is present in the ER, Golgi and plasma membranes (282, 398). It has an N-terminus transmembrane domain that was reported to form ion-channel on the cell membranes (109). Function of this ion-channel is not clear but might be involved in virus release (339). It is known that interferon-\(\alpha\) activates various antiviral genes in the infected cells, some resulting in the retention of virus particle on the cell surface (274). It was discovered that an interferon-\(\alpha\) induced protein, B cell stromal factor 2 (BST-2), is responsible for this virus retention property (275). BST-2 is also known as CD317 or tetherin. Importantly, Vpu can bind and target BST-2 for degradation, thereby annulling the inhibitory effect of tetherin (182, 275). Vpu can
also mediates the down regulation of cell receptor CD4 in the ER (411). This was achieved with help from the human beta-transducene repeat containing protein (βTrCP), which act as the connector protein between Vpu and the cytoplasmic domain of CD4 (247). Binding of Vpu to βTrCP and CD4 eventually caused the degradation of CD4 in the cytosolic proteasome (28). Some studies also suggest that Vpu might down regulate the presentation of major histocompatibility complex class I (MHC-I) and II (MHC-II) (175, 197).

Nef is a 27 kDa myristolyated protein expressed by all HIV-1 and SIV. Its importance in viral pathogenesis was clearly demonstrated when patients infected with Nef defective HIV-1, did not progress to AIDS rapidly (89). Nef is known to down regulate the expression level of cellular receptor CD4. This occurs through binding to the cytoplasmic tail of CD4, internalization via the clathrin coated pits and redirecting them to lysosome for degradation (62, 410). There is evidence to suggest that Nef may have a similar effect on CCR5 and CXCR4 expression as well (169, 260). By reducing the expression level of these cellular receptors, Nef can prevent multiple HIV-1 virions from infecting the same cell, which could affect viral replication (198). Additionally, Nef decreases the MHC-I expression level on cell surface by redirecting the newly synthesized MHC-I to the trans-Golgi network and eventual degradation in the lysosome (343). A study has suggested that Nef may have similar effect on MHC-II but the mechanism is not clear (337). These down regulations may result in less viral antigen presentation to the immune cells, such as CD8+ T-cell, thereby protecting the infected cells from immune surveillance. However, the detection of escape mutants from CTL responses implies that the down regulation of MHC-I or II may not be efficient (144).
Interestingly, some studies had implicated Nef in blocking the tetherin activity in a species-specific manner. They reported that SIV Nef has the ability to antagonize tetherin in non-human primates but not in human. On the other hand, HIV-1 Nef has no such functions and human tetherin can only be inhibited by HIV-1 Vpu (187, 437).

**Tat and Rev**

Tat is a ~15 kDa transactivator protein that is synthesized after the mRNA splicing of its two exons, located upstream and overlaps with the envelope gene. Tat contains two domains that bind to the TAR in vRNA and recruit cellular transcriptional machinery (177). Surprisingly, these domains are situated within the exon 1, making exon 2 essentially non-functional (213). Moreover, the presence of a nuclear localization signal within Tat allows its accumulation within the cell nucleus (155). The main function of Tat is to bind to the RNA stem-loop structure of TAR for transcription initiation and elongation of the viral transcripts. In the absence of Tat, viral transcripts tend to terminate prematurely (190). Also, Tat was reported to affect the expression of several cellular genes such as cytokines and MHC (253, 334). Moreover, a number of studies have suggested that Tat can be secreted from the infected cell and entered adjacent naïve cells, resulting in the induction of apoptosis and cell death (58, 134). Together, these functions of Tat may contribute to viral pathogenesis.

Rev is a ~19 kDa protein that is encoded by two exons positioned upstream and within the envelope gene. Rev primarily served to export the unspliced and singly spliced vRNA from the nucleus into cytoplasm. It achieved this function by binding to the Rev response element (RRE), located within the envelope gene, which is carried by all
unspliced and singly spliced vRNA (243, 245). The interaction domain with RRE has been identified as an arginine-rich RNA binding motif positioned in the exon 2 of Rev. This RNA binding motif also double as the nuclear localization signal that allows Rev entry into the nucleus (386). The C-terminus of Rev also contains a leucine-rich motif that functions as the nuclear export signal, allowing export of the Rev-RRE complexes (407). Multimerization of Rev onto the RRE is required for activation of this nuclear export ability (183, 380). This ability of Rev to nuclear export the unspliced vRNA allows the incorporation of full-length viral genome into the nucleocapsid during virion assembly. A detailed crystal structure of Rev has been recently resolved (83, 97).

**Envelope**

Envelope is notably the most studied among all the HIV-1 proteins, due to its indispensable functions in multiple aspects of the viral life cycle and *in vivo* survival. The *env* is translated in the ER as a ~160 kDa precursor glycoprotein known as gp160. During its synthesis process, the gp160 is glycosylated, trimerized and later cleaved into gp120 and gp41 by the cellular enzyme furin (129, 268). As described earlier, gp120 and gp41 are also designated as the surface (SU) and transmembrane subunit (TM) of the envelope, respectively. The SU encompasses the receptor binding domains, while the TM promotes viral envelope fusion with the target cell membrane after receptors binding via its fusion domain, as well as envelope incorporation into the virion via its cytoplasmic tail as described earlier (25, 123, 296, 424). Detailed structure of the HIV-1 envelope has been published (246). This review will center on the SU given its dominant role in viral pathogenesis and it is the focus of this dissertation.
SU has the greatest genetic diversity amid various HIV-1 strains, whereas TM is highly conserved. Based on studies with clade B HIV-1, the distribution of this genetic diversity is not uniform in the SU, but is located mainly within its five hyper-variable domains, V1 to 5, which are interspersed by five constant regions, C1 to 5. However, data from this dissertation suggests that for clade C HIV-1, certain portion of the constant regions, such as the N-terminus of C3, might also accommodate higher genetic diversity than generally anticipated. Among the variable domains, V1-V2 is the most polymorphic and can tolerate extensive deletions, insertions as well as various patterns of N-glycosylation sites (211). Studies have demonstrated that V1-V2 can influence viral tropism and shield the neutralization sensitive epitopes that lie within the V3 loop from host humoral responses (206, 207, 215, 293, 423). Moreover, a study on SIV has shown that V1-V2 will undergo structural changes to facilitate the formation of co-receptor binding sites for CCR5 or CXCR4, after the viral envelope binds to its primary receptor CD4 (61). Others have documented the V3 loop as the main determining factor in viral co-receptor usage (178). For instance, mutations in charged residues within V3 are associated with switching co-receptor usage (86, 163, 181, 355). Similarly, insertions or deletions within V3 can also influence the viral co-receptor usage (441). Together, these studies underscored the importance of V1, V2 and V3 on the viral co-receptor usage. Additionally, mutations within V2 and V3 can impact the membrane fusion process as well (114, 150, 283). Despite the importance of V1, V2 and V3 in co-receptor usage, deletion analyses have demonstrated that truncated SU lacking these regions can still bind to the primary receptor CD4 (297, 425). Unlike these variable regions, which do not play a direct role in CD4 binding, some constant regions within the SU were identified as
participating directly in the binding of CD4 (215). Surprisingly, the functions of V4 and V5 are relatively less established. Although these two variable regions are not implicated in receptor or co-receptor binding, their presence may preserve the structural integrity of the envelope, since deletion of V4 may influence processing of the envelope (425).

In addition to determining the viral tropism, SU functions as the principal viral defense mechanism against the potent host humoral responses. SU achieved this ability through masking of its potential neutralization sensitive epitopes from the host immune surveillance by camouflaging itself extensively with carbohydrates derived from the infected cell. This resulted in a heavily glycosylated envelope protein, with over 50% of its molecular weight derived from carbohydrates (133). The addition of carbohydrates to protein is known as glycosylation. The HIV-1 envelope is N-glycosylated at the consensus sequences NXS or NXT as it traffics through the Golgi apparatus to the cell membrane. Amino acid residues that are N-glycosylated are known as potential N-glycosylation sites (PNGS) and the majority of them are located within the SU. Just like the genetic diversity, PNGS are unevenly distributed within the SU and highly variable within the V1-V5 region. The hyper-variable nature of the V1-V5 region provided additional capital for the envelope to maintain a constantly changing number and pattern of PNGS on the virus surface. This pattern of PNGS was proposed to function as a “glycan shield” that evolves in response to the host immune surveillance and shields the neutralization sensitive epitopes of the envelope from host humoral response (405). Evidence supporting this glycan shield theory came from SIV mutants lacking certain PNGS in its envelope elicited higher antibody neutralization activity in the infected monkeys (309). Moreover, it is well established that changes in the PNGS pattern on the
HIV-1 envelope will influence the neutralization susceptibility of the virus (18, 128, 145). PNGS is also implicated in the conformation and structural integrity of the HIV-1 envelope (224, 422).

Lastly, the HIV-1 envelope has been shown to down regulate the CD4 expression level by binding to and retaining newly synthesized CD4 in the ER (220). Envelope may also induce cytotoxicity in infected cells by inducing cytokine production (431). Importantly, HIV-1 envelope can also induce apoptosis of infected CD4+ T cells (222). This may contribute to by-stander T cell death and overall decline of CD4+ T-cell counts in AIDS patients. Given all these critical roles, tremendous effort has been invested in vaccine as well as therapeutic strategies targeting the HIV-1 envelope. There are currently only two classes of drugs available against the HIV-1 envelope; maraviroc that antagonizes the co-receptor CCR5 and enfuvirtide that inhibits the viral envelope from fusing to the cell membrane (99, 216).

**Pathogenesis of HIV-1**

**Life cycle of HIV-1**

In order to better understand HIV-1 pathogenesis, it is important to briefly review the HIV-1 replicative cycle, which is separated into several steps. The first step is the binding of viral Env to their primary receptor, CD4 molecules, expressed on the host cell surface. Binding of primary receptor triggers a conformational change to the SU and exposes the viral co-receptor binding sites (140). The binding of viral Env to its co-receptor, either the CCR5 or CXCR4 chemokine receptor depending on the tropism of the respective virus strain, activates additional conformational changes in the TM that
permits fusion of the virus-host cell membrane and releases the viral core into the cytoplasm (258). Subsequently, the vRNA will undergo reverse transcription to complete the synthesis of the double stranded viral DNA and follow by the uncoating of viral core. Formation of the PIC will transport the newly synthesized vDNA into the nucleus, which will then integrate into the host chromosome to form the provirus. This process can be completed within 36-48 hrs post-infection (400). Many host proteins are involved in these early replication steps (115). For example, cyclophilin A is necessary for the uncoating of viral core (231).

The generation of new virions begins with transcription of the provirus and translation of viral genes using the host cellular machineries. The Tat protein enhances the transcription of viral genes by binding to TAR as described earlier. This process yield full-length viral transcripts, some are alternatively spliced by the cellular mRNA splicing machineries into smaller viral transcripts. As described earlier, some viral proteins are encoded in separate exons. The viral proteins Tat, Rev and Nef are translated from doubly spliced viral mRNA; while Vpu, Vpr, Vif and Env are translated from singly spliced viral mRNA. The major structural protein Gag-Pol is translated directly from the unspliced full-length viral transcript. Since the splice sites of HIV-1 are not efficiently utilized, some full-length unspliced viral mRNA will be retained and serve as the new viral template. The unspliced and multiply spliced viral mRNA are then transported by Rev into the cytoplasm for translation. The viral structural proteins, such as the Gag-Pol precursor polypeptide, are produced first and are the most abundant. These structural proteins are trafficked to the site of assembly, namely the plasma membrane, and initiate virion assembly (280). The new viral templates are incorporated into the nucleocapsid.
Finally, the fully assembled virion buds from the host cell as an immature and noninfectious virion via the ESCRT pathway as described earlier (390). The subsequent proteolytic cleavage of Gag by PR triggers the morphological maturation of the HIV-1 and form the infectious virion.

**Transmission**

Transmission is the first step in HIV-1 pathogenesis and can occur by multiple routes. HIV-1 can be transmitted by exposure to blood or blood-related products from infected individuals. Intravenous drug users have the highest risk for infection by this transmission route. Over 50% of intravenous drug users are infected with HIV-1 in developing countries, for example Estonia and Mauritius. Even in developed countries, Canada for example, HIV-1 prevalence rate is near 6% among intravenous drug users (3). Nearly 80% of HIV-1 transmissions in adults are due to unprotected sexual contact with infected individuals via mucosal routes, such as rectal or vagina (2). The mucosal route is the most common among homosexual and heterosexual couples. It was estimated that homosexual couples have 10 fold higher infections risk than heterosexual couples, which could be due to physiological differences between the rectum and vagina (375). Since these two routes involve transmission of the virus between non-mother-child relationship individuals, they are classified as horizontal transmission. HIV-1 can also be transmitted vertically from the mother to her child, which can occur during pregnancy, delivery or breast-feeding. Nearly 0.4 million children were infected by HIV-1 through mother-to-child transmission in 2009 (2). Although the exact mechanism is unclear, the fetus appears to be infected by the virus that crosses the placenta barrier in-utero or the newborn may come in contact with infected maternal blood during delivery or via breast
milk after delivery. It is known that HIV-1 infected cells and cell-free viruses are present in breast milk with vRNA concentrations ranging from 240 to over 900 copies per ml (228). This highlights the importance to stop breast-feeding in HIV-1 positive mothers. The use of formula for newborns together with administration of antiretroviral therapy drugs to HIV-1 infected pregnant women is highly effective in preventing perinatal transmission; optimistically, it is anticipated that mother-to-child transmission might be eliminated in the near future (3, 60, 368).

Despite these various transmission routes, HIV-1 has low transmissibility. For example, successful heterosexual transmission only occurs after numerous extensive exposures to the virus, as shown by studies on commercial sex workers (100, 112, 404). Also, only an estimated 30% of the children born to HIV-1 positive mothers are infected in the absence of antiretroviral therapy (10, 85). Some data even implies that exposure to infected blood products may not result in HIV-1 infection (209). Together, these reports suggest that HIV-1 has a low efficiency in establishing infection during transmission. Many factors such as the virus concentration of the donor, integrity of the mucosal membrane, chronic inflammation, defective viral genes and other host genetic factors could all determine the transmissibility of HIV-1. Studies show that individuals with a homozygous 32 nucleotide deletion in their CCR5 gene display remarkable resistance to HIV-1 infection (236). The importance of this gene mutation in HIV-1 transmissibility is best demonstrated when a HIV-1 patient, who has high viral load during chemotherapy, maintained an undetectable viral load without antiretroviral therapy after receiving stem-cell transplant from a CCR5-delta-32 donor (176). This mutation is present in approximately 15% of the Caucasian population but rare in Africans and Asians (328).
Even after the initial transmission event, HIV-1 still has a number of significant barriers towards establishing an infection. Following an initial infection, mucosal transmission is usually established by a small founder population, which suggests a bottleneck exists for the invading HIV-1 (6, 143, 196, 326). Moreover, viruses with certain phenotypes, such as fewer PNGS and neutralization resistance towards the maternal antibodies are preferentially selected to cross this bottleneck (93, 417, 420). Importantly, this genetic bottleneck also occurs during SIV mucosal transmission models and is correlated with low-dose, but not high-dose inoculum, which mimics transmission in humans (262, 350, 364). The fact that this bottleneck was not observed in blood-borne transmission suggests the mucosal components, such as cell types and membranes, may be accountable for the genetic bottleneck (22). There are still controversies surrounding the transmitted virus phenotypes with no definitive conclusions. Many groups, including our own, are actively investigating this issue.

**Infection**

After the transmission event, viruses that had overcome the mucosal barriers will come in contact with their target cells and commence the infection cycle. Importantly, regardless of the transmission route, HIV-1 tends to converge on the gut-associated lymphoid tissues (GALT) of the gastrointestinal tract as it contains the majority of the CD4$^+$ lymphocyte that are susceptible to HIV-1 infection (269). Targeting the gut tissue was reportedly due to the specific affinity of HIV-1 gp120 for the gut homing receptor $\alpha 4\beta 7$, which are expressed on gut associated CD4$^+$ T cells (73). Although dendritic cells cannot be productively infected by HIV-1 due to the host restriction factor SAMHD-1, Langerhans’ cells in these tissues can facilitate the spread of HIV-1 by transporting the
virus to uninfected T-cells (20, 218). As mentioned earlier, CD4 is the primary receptor for HIV-1 and can be found on T-lymphocytes, macrophages and certain populations of dendritic cells. However, it is not indispensable as some HIV-1 isolates are CD4-independent or use CD8 as their primary receptor, but those are rare exceptions (323, 324, 435, 436). HIV-1 infection also requires the presence of co-receptors CCR5 or CXCR4, which can be found on macrophages and T-lymphocytes. Virus isolates that use CXCR4 are exclusively T-lymphocytes tropic and form syncytium, while those that use CCR5 can infect macrophages as well as T-lymphocytes and non-syncytium inducing. Primary viral isolates from early transmission are predominantly CCR5 tropic (236). On the contrary, the emergence of CXCR4 tropic and CXCR4/CCR5 dual tropic viruses are associated with an advanced course of disease and faster progression to AIDS (49, 200, 214). Additionally, CXCR4 tropic viruses replicate faster than CCR5 tropic variants in culture (16, 64, 376, 391). It is still not fully understood why the virus switches its co-receptor tropism from CCR5 during early infection to CXCR4 during late infection. Presumably, the availability of CCR5 target cells decreases at the late course of disease and hence selects for the expansion of CXCR4 tropic viruses. It has been suggested that gp120 may evolve into a more open structure, which could accommodate the CXCR4 structure upon disease progression.

After locating the target cell, the virus binds and enters host cells, as described earlier, and eventually is integrated into the host chromosomes. Once integrated, the virus can undergo replication or remain latent. Latency is the most critical issue for all HIV-1 therapies. During latency, very little or no viral proteins are expressed. This provides a vital advantage for the virus by allowing it to remain undetectable by the host immune
surveillance. Latent HIV-1 can be found in resting memory CD4+ T cells as well as hematopoietic progenitor cells (HPCs). However, data on HPCs is still controversial (52, 72). These latent HIV-1 reservoirs can remain stable for extended period of time, making the eradication of HIV-1 impossible with current treatment strategies (356, 365).

Moreover, viruses with drug resistance mutations that evolved in patients on HAART might enter latency and reactivate during therapy interruption, contributing to the spread of drug resistant HIV-1 strains (188). The exact mechanism that regulates HIV-1 latency is still unclear. Given that HIV-1 is integrated into the host genome, it is subjected to the same regulatory mechanisms, such as DNA methylation and histone acetylation, as the host. Recent studies also suggest that additional TFBS within the HIV-1 LTR promoter region enhance the establishment of latency (104). Since the current antiretroviral drugs cannot eradicate latent HIV-1, new classes of drugs are being actively pursued to purge the latent reservoirs in order to achieve a cure of HIV-1 infection. For example, prostratin can reactivate latent HIV-1 in the presence of HAART (212). However, drug toxicity and reactivation of other latent viruses, such as herpesviruses, remain the major obstacles for the development of these drugs (40).

**Disease progression**

Following the infection of GALT, viruses that did not enter latency begin replication and produce localized infections involving relatively small number of cells. Studies from the SIV model demonstrated this process occurs near the site of entry within 3 days of initial infection (262). This stage is known as the eclipse phase of disease progression and might last for an average of 10 days (Fig 2A) (233). During this stage, the viral load is usually below the detection limit and has no detectable HIV-1 specific
host immune responses. Without any host immune constraints, viral replication continues to generate new virions at these localized infection sites, resulting in massive losses of gut-associated CD4+ T cell (399). Eventually, the HIV-1 infected cells and virions begin to migrate towards the lymph nodes, possibly with assistance from dendritic cells as mentioned earlier.

Through the lymphatic network, HIV-1 disseminates systematically and the viral replication increases exponentially as it gains access to other CD4+ T cells in the body. This resulted in detectable viral load in the peripheral blood and the start of the acute phase of infection. This acute phase takes place ~30 days post-infection and is marked by a peak viremia of ~10^6 copies/ml of vRNA from blood (Fig 2A). It is also accompanied by decreasing (but still within the normal range) of CD4+ T-cell counts (Fig 2B). The viral load and CD4+ T-cell counts are the standard benchmarks for determining disease progression in HIV-1 infections. When the CD4+ T-cell count falls to <200 cells/ul, the patient is considered to have AIDS as defined by the CDC (338). The high viremia during the acute phase also elicits host immune responses. Acute phase often coincide with the surge of inflammatory cytokines such as interferon-α (359). In addition, natural-killer (NK) cells are also activated (9). Simultaneously, activation of CD8+ T cells denotes the beginning of cytotoxic T-lymphocyte (CTL) responses (142). However, the effectiveness of CTL responses is dependent on the human leukocyte antigen (HLA) type of the infected individual. Some HLA types, such as HLA-B*27 and 57, are associated with protection against HIV-1, whereas others such as HLA-B*35 and 51 are associated with disease progression (50, 51, 194). The increase in viral antigens also elicits the host humoral responses. However, the antibodies produced at this stage are usually non-
neutralizing (382). Due to immune activation, some infected individuals might experience flu-like symptoms or others such as body rashes and headache. These symptoms are usually short lived and resolved within a few weeks.

The activated immune system reduces the viral load to a significantly lower level and keeps it stable over extended periods of time. This establishes the viral set point and starts the chronic phase of infection (Fig 2A). Studies have also correlated a higher viral set point with rapid disease progression (259). The chronic phase of infection can last for years without any significant disease symptoms. It is during this period that the host immune system goes into over drive, as it continuously tries to detect and eliminate cell-free viruses or HIV-1 infected cells. The host humoral response produces abundant non-neutralizing antibodies against HIV-1. However, some studies suggest that non-neutralizing antibodies might still exert significant pressure on the virus by inducing antibody-dependent cell-mediated cytotoxicity (ADCC) (166, 388). Importantly, HIV-1 specific neutralizing antibodies first appear after chronic exposure to viral antigens (98, 332). This is due to the time between antigen recognition and antibody affinity maturation. Neutralizing antibodies primarily target the HIV-1 envelope and are predominantly IgGs (30). It is difficult to elicit high level of broadly neutralizing antibodies that recognize various HIV-1 strains and are found in ~1% of infected individuals (357). In response to the humoral response, the viral envelope evolves mutations, for example changes in PNGS or length of their variable loops, to escape neutralization. Recently, a tier system was proposed to classify the neutralization sensitivity of HIV-1 envelopes, with tier 1 being the most neutralization sensitive (344). Beside the humoral response, CTL is also active during the chronic stage and contributes
significantly towards maintaining a stable viral load, as shown by SIV experiments with depleted CD8+ T cells (116).

Despite the various immune responses, CD4+ T-cell counts continue to decline throughout the chronic phase and the patient enters the AIDS phase as they fall to <200 cells/ul (Fig 2A). At this point, chronic immune activation has exhausted most of the CD4+ T cell populations. Consequently, viremia rebounds sharply and the collapse of the immune system occurs. Furthermore, opportunistic infections, such as PCP, begin to develop in these patients. In rare cases, infected patients do not progress to AIDS for over 10 years. These patients have low to undetectable viral load and remain healthy without any antiretroviral treatments. These individuals are known as long-term non-progressors (LTNP) or elite controllers, opposed to rapid progressors who progressed to AIDS. Host genetic factors, such as HLA type and CCR5-delta-32, are believed to be critical in these cases. Detailed functional changes to the HIV-1 envelope during the entire disease progression are not well established in general. Most studies only focused on the acute phase or the neutralizing characteristics of the envelopes. Data from this dissertation will contribute towards our understanding of this issue.

**Non-human primate models**

**SIV and SHIV models**

Non-human primate models are an invaluable resource for many biomedical studies. It has broadened our understanding of numerous infectious diseases and facilitated the developments of life-saving vaccines, drugs and therapeutic procedures. Since the discovery of SIV infection in macaques, non-human primates have been an
important model for modeling HIV-1 disease progression. Among the various non-human primate species, rhesus macaques (*Macaca mulatta*), cynomolgus macaques (*Macaca fascicularis*) and pigtailed macaques (*Macaca nemestrina*) are most commonly used. They share similar physiology, anatomy, genes and immune system as human. Macaque’s close phylogenetic relationships with humans also made them an ideal human surrogate. Although HIV-1 is derived from chimpanzee, it is a highly endangered species and not a viable research model. On the other hand, macaques are readily available as they can be bred in captivity for research purposes. Macaque’s relatively short life span, compared to human, allows studying of disease progression in a reasonable time frame, which might otherwise takes years or decades to develop. Because of its vital role in research, tremendous resources were invested to sequence the complete genome of rhesus macaques, which is now available for researchers (136).

Non-human primates are the natural host of SIV. Surprisingly, most naturally infected monkeys do not progress to AIDS despite high plasma viral load (11, 312). The abilities for naturally infected monkeys to rapidly control their immune activation after infection, reduces the chronic immune activation and preservation of immune cells, marks some of its differences from animals experimentally infected with SIV from other species (36, 186). The precise factors involved are not fully understood, but many hypotheses have been proposed (57). Macaques will only progress to AIDS after infection by SIV from another species (225). Importantly, the infected monkeys experience massive loss of CD4+ T cells, high viral load, susceptible to opportunistic infections and eventual progression to AIDS; comparable to the HIV-1 disease progression in human (232, 399, 403). Similar to human, some macaques have monkey
leukocyte antigen (MLA) types, such as Mamu-A*01, Mamu-B*08 and 17, that are associated with slower disease progression (159, 237). SIV shares similar genomic structure, viral proteins, and functions when compared to HIV-1. Moreover, SIV are predominantly CCR5 usage, as with the majority of primary HIV-1 isolates, and target the same susceptible cell populations as HIV-1 (63, 199, 443). To avoid host immune suppressions, SIV utilizes the same evasion mechanisms as HIV-1, such as the modification of envelope PNGS pattern and mutations of Gag CTL epitopes (56, 144). These similarities make SIV an ideal model for understanding disease progression in HIV-1 infected patients. Furthermore, homogenous SIV population can be used to infect the macaques under a controlled environment via any choice of transmission routes and dosages. The infected animals can be monitored closely and samples collected frequently. All these variable experimental conditions and procedures are impossible to perform on humans. By using this animal model, we have expanded our knowledge of HIV-1 from the viral protein functions to disease transmission and progression (255, 261, 288, 408).

For example, the SIV infection model has contributed the most towards our understanding of viral escape from neutralizing antibodies and CTL responses during infection (106, 144, 418). Additionally, this animal model has broad applications with respect to developing an effective HIV-1 vaccine (284, 285). Recently, the SIV infection model was used to mimic the RV-44 Thai HIV vaccine trial study, in hopes of understanding the limited protective results observed in that clinical trial (289).

Regardless of the close similarities between HIV-1 and SIV, their envelopes possess several important conformational differences (286). This is supported by reports showing that neutralizing antibodies against HIV-1 cannot cross-neutralize SIV (434). To
overcome these issues, the *tat, rev, vpu* and *env* genes of HIV-1 were inserted into the molecular clone SIVmac239 to create the chimeric virus, SHIV (353). Unfortunately, the early versions of SHIV are non-pathogenic in animals (226, 308). The first pathogenic SHIV, SHIV-89.6P, which causes AIDS in the infected animals, was achieved after *in-vivo* passages in several macaques (307). Using this SHIV model, the protective effect of passively transferred antibodies against HIV-1 was first demonstrated (429). It provides evidence that humoral responses can be used to develop an anti-HIV-1 vaccine. Others have utilized this model to examine the HIV-1 envelope evolution during disease progression (31). Despite the success of these SHIVs, there were concerns that the rapid course of disease after infection may not represent the normal HIV-1 disease progression in human (307). Moreover, these SHIVs were generated with lab-adapted HIV-1 strains, which are clade B CXCR4 tropic and do not reflect the CCR5 tropic primary isolates (307). To this end, new SHIVs that express clade B CCR5 tropic envelopes were generated (152, 276). To further improve the existing model, our collaborators generated a R5 SHIV-C that expresses a CCR5 tropic clade C primary isolate envelope from a Zambian infant (172, 358). This SHIV-C model is the first clade C SHIV that was pathogenic in monkeys and infected animals progressed to AIDS. These new and improved SHIV models are now employed in various vaccines and mucosal transmissions studies (131, 219, 272).

**Simian-tropic human immunodeficiency virus type 1 (stHIV-1) model**

SIV and SHIV infections of non-human primates are beneficial models for HIV-1 research. However, the lack of important HIV-1 genes, such as *gag* and *pol*, created some intrinsic disadvantages for these models. The lack of HIV-1 *gag* gene
renders these models unsuitable for testing anti-HIV-1 vaccines that elicit CTL responses, which mainly target the Gag protein. In addition, certain antiretroviral drugs, such as NNRTI, have no effect on the SIV RT (84). A new animal model utilizing the simian-tropic HIV-1 (stHIV-1) infection of pigtailed macaques was proposed to address these issues (153, 189). Unlike SHIV, stHIV-1 maintains the majority of HIV-1 genes with the exception of \textit{vif}, which is replaced by its SIV equivalent.

HIV-1 is non-pathogenic in non-human primates due to the host restriction factors, APOBEC3G and TRIM5α, present in the monkeys (127). Both factors must be overcome for successful infection of monkeys by the stHIV-1 (244, 366). As described earlier, the activities of APOBEC3G is highly species specific, where the monkey APOBEC3G can only be inhibited by the SIV Vif (33). Thus, the replacement of HIV-1 Vif by SIV Vif is needed to overcome the monkey APOBEC3G restriction and allow stHIV-1 infection of monkey cell lines (154, 180). TRIM5α is a member of the tripartite motif (TRIM) family of proteins that consists of the RING, B-Box, Coiled-Coil and SPRY domains. Although its normal cellular function is unclear, TRIM5α was implicated in the regulation of cell cycle as well as apoptosis (201, 389). Additionally, evidence suggest that the primary function of TRIM5α is mounting antiviral responses, as its expression level is interferon inducible (15). Importantly, the restriction effect of TRIM5α is also species-specific as the monkey TRIM5α inhibits HIV-1 but not SIV and vice versa in human (195, 366). TRIM5α may inhibit viral infection by accelerating the uncoating of the viral core in the cytoplasm (367), thereby, disrupting the reverse transcription process and preventing the formation of vDNA. Interestingly, pigtailed macaques express two non-functional TRIM5α isoforms that render it susceptible to
HIV-1 infection (38). Hence, stHIV-1 overcomes the APOBEC3G and TRIM5α restrictions by using the pigtailed macaques as its host animal. Evidence from in-vitro studies showed that stHIV-1 can infect pigtailed macaque PBMC and maintain similar replication kinetics as SIV (377).

Unfortunately, the success of in-vitro stHIV-1 infection was not translated into disease progression in-vivo (153, 377). Despite the detection of viral DNA in the PBMC, the infected animals displayed no significant decrease in CD4+ T-cell count and co-culture with uninfected donor macaque PBMC was unable to rescue any viable viruses (377). These observations confirmed the establishment of infection by stHIV-1, but underscored its inability to induce AIDS in pigtailed macaques. These discrepancies between the in-vitro and in-vivo data suggest that certain HIV proteins, such as Nef, may not be adequate to maintain pathogenicity in monkeys and needs to be replaced by their SIV counterpart. More research will be needed to explain this low pathogenicity of stHIV-1, and stHIV-1 needs to be optimized before it can be used as an animal model for HIV-1 infection of human.

**Research aims**

The HIV-1 envelope is the major target for developing an anti-HIV-1 vaccine. However, the high viral diversity presents a major problem for designing an effective vaccine. An effective vaccine must stimulate immune responses against a broad range of viral populations, including viruses that will evolve during disease progression. Therefore, a better understanding of how the HIV-1 envelope V1-V5 region evolved during disease progression is critical. Unfortunately, it is nearly impossible to study this
issue in humans due to numerous logistical and ethical constraints. Hence, an animal model that closely mirrors HIV-1 disease progression in humans is needed. Additionally, this animal model needs to allow the infection to be established by a homologous virus population under a controlled environment and infected animals must be monitored over the entire disease course. Moreover, HIV-1 clade C is the predominant strain in sub-Saharan Africa where over 70% of the HIV-1 infected patients reside. Thus, it will be important to study and characterize the HIV-1 clade C envelope.

The main objective of this dissertation was to investigate the evolution of HIV-1 clade C envelope V1-V5 region during disease progression. Our approach was to infect non-human primates with SHIV, which expresses the gp120 of a HIV-1 clade C primary isolate derived from the infected Zambia infant 1157i, and monitor the infected animals throughout the course of disease.

Specific aim one was to establish the validity of this non-human primate model by comparing the longitudinal genetic variations of the envelope clones, V1-V5 region, derived from the infected animals and patient over the disease course, and correlate the observed genetic mutations with disease progression. Specific aim two was to determine the longitudinal changes in the biological functions of these envelope clones at various disease stages and correlation with disease progression.

We hypothesized that the infection of non-human primate with SHIV-C would resemble the HIV-1 disease progression in patient. The genetic evolution of the envelope V1-V5 region that occurs in the infected animals may mirror those observed in patient and correlate with disease progression. In addition, the envelope clones derived from the
later stages of disease may possess different biological characteristics, such as faster kinetics or better fusion capability, compared to those present at earlier disease stages.
Figures and Figure legends

Figure 1.

A. HIV-1

B. SIV

C. SHIV
Figure 1. Schematics of the genomic structure of (A) HIV-1 (B) SIV and (C) SHIV. Each rectangular bar represents an encoded protein, with the exception of 5’ and 3’ LTR which are non-coding regulatory regions. Yellow indicate genes derived from HIV-1. Diagrams are not drawn to scale. The entire genome is generally ~9.5 kb nucleotide in length.
Figure 2.

A. Plasma viral RNA (copies/ml) as a function of time post infection.

B. CD4+ T-cell count (cells/µl) as a function of time post infection.
Figure 2. Different disease stages of HIV-1 infection. (A) Plasma viral RNA load (B) CD4+ T-cell count. The days or years post-infection presented are estimates.
Chapter 2

A comparative study of HIV-1 clade C env evolution in a Zambian infant with an infected rhesus macaque during disease progression

For Yue Tso\textsuperscript{a}, Federico G. Hoffmann\textsuperscript{a,\dagger}, Damien C. Tully\textsuperscript{a}, Philippe Lemey\textsuperscript{d}, Robert A. Rasmussen\textsuperscript{b,c}, Hong Zhang\textsuperscript{a}, Ruth M. Ruprecht\textsuperscript{b,c} and Charles Wood\textsuperscript{a*}

\textsuperscript{a}Nebraska Center for Virology and the School of Biological Sciences, University of Nebraska-Lincoln, Nebraska 68588; \textsuperscript{b}Dana-Farber Cancer Institute and \textsuperscript{c}Harvard Medical School, Boston, Massachusetts, 02115; \textsuperscript{d}Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium

*Corresponding author

Mailing address: Nebraska Center for Virology, School of Biological Sciences, University of Nebraska-Lincoln, Morrison Center P.O. Box 830666, Lincoln, NE 68583-0900

Phone: (402) 472-4550; Fax: (402) 472-3323; E-mail: cwood@unlnotes.unl.edu

\dagger Present address: Instituto Carlos Chagas-FIOCRUZ, Curitiba, Paraná, Brazil.

Data from this chapter was published in AIDS 2009, 23:1817–1828
Abstract

**Objective:** To evaluate whether HIV-1 clade C (HIV-C) envelope variations that arise during disease progression in rhesus macaque model reflect changes that occur naturally in human infection.

**Design:** An infant macaque was infected with SHIV-1157i, an R5 tropic clade C SHIV, that expresses a primary HIV-C envelope derived from an infected human infant and monitored over a 5-year period. Genetic variation of the V1–V5 envelope region, which is the main target for humoral immune responses, derived from the infected macaque and infant was examined.

**Methods:** The V1–V5 envelope region was cloned and sequenced from longitudinal peripheral blood mononuclear cell samples collected from the infected macaque and infant. Phylogenetic analysis [phylogenetic tree, diversity, divergence, ratio of nonsynonymous (dN) and synonymous substitution (dS) and dN distribution] was performed. Plasma RNA viral load, CD4⁺ T-cell count, changes in the length of V1–V5 region, putative N-linked glycosylation site number and distribution were also measured.

**Results:** Phylogenetic analysis revealed that changes in the macaque closely reflected those of the infant during disease progression. Similar distribution patterns of dN and hot spots were observed between the macaque and infant. Analysis of putative N-linked glycosylation sites revealed several common variations between the virus populations in the two host species. These variations correlate with decline of CD4 T-cell count in the macaque and might be linked with disease progression.
**Conclusion:** SHIV-C infection of macaque is a relevant animal model for studying variation of primary HIV-C envelope during disease progression and could be used to analyze the selection pressures that are associated with those changes.
**Introduction**

The V1–V5 region of HIV-1 envelope glycoprotein is the focus of numerous studies because of its critical role in viral pathogenesis and immune evasion (130, 133, 322, 333, 362, 379, 405). Because of its hypervariable nature, the majority of the potential vaccines targeting the envelope have thus far failed to elicit sterilizing immunity against heterologous viral challenge in animal studies and human clinical trials (111, 137, 147, 251, 294, 305, 428). Thus, it is important to better understand the changes of the V1–V5 region that occur during the course of disease and the factors that contribute to these changes.

Non-human primate models that utilize SHIV (156, 192, 240, 308, 354, 401) had been used to study clade B and laboratory-adapted strain envelope variants during the course of disease (31, 165, 208). It has been suggested that there is a close similarity in the envelope evolution during SHIV infection of nonhuman primates and HIV-1 infection of humans; similar envelope mutations were observed over time in a laboratory worker accidentally infected with HIV-1 IIIB and a macaque experimentally infected with a SHIV expressing the identical envelope (24, 165). However, HIV-1 IIIB is an X4 laboratory-adapted clade B virus whose envelope evolution may not be representative of that of R5 tropic primary isolates, and only a small number of sequences from both the macaque and human were analyzed. Several studies have also indicated that the env of different HIV-1 clades can evolve differently under selective pressure (71, 130, 203). To adequately determine if SHIV infection of macaques is a viable model for assessing envelope evolution, SHIV constructs using env from recently transmitted and
biologically relevant primary HIV-1 isolates must be tested. The evolutionary biology of the R5 SHIV-C in this current study, SHIV-1157i, represents such a virus.

We have been following a cohort of mostly HIV-C infected mother/infant pairs in Lusaka, Zambia. Among them, infant 1157i, a slow progressor, was followed prospectively, and viruses derived longitudinally from the infant have been extensively characterized (438). SHIV-1157i was generated with an R5 tropic env derived from infant 1157i at the age of 6 months. An infant macaque was infected with SHIV-1157i and monitored through disease progression until it died from AIDS after approximately 5 years (172, 358). Over this period, we tracked the changes of the V1–V5 region from the time of inoculation to euthanasia of the infected macaque. Together with data previously gathered for infant 1157i, we were able to compare the genetic variation of the V1–V5 region of this particular primary clade C isolate during disease progression in both species (438).

As our study utilized a clade C primary R5 tropic env instead of an X4 laboratory-adapted clade B strain and we analyzed over 300 viral envelope sequences, our investigation represents a more in-depth comparison of viral evolution during disease progression. This study has provided a unique opportunity to identify potential envelope mutations in the infected macaque, which may associate with and predict future disease progression in the infected infant.
Materials and methods

Construction of SHIV-1157i

SHIV-1157i contains env of a primary HIV-C isolate from a 6-month-old Zambian infant 1157i. PvuI (P) was introduced into the 3’ half of SHIV-vpu⁺ proviral DNA. The 2.0-kb KpnI (K)-PvuI fragment of HIV1157i was amplified to replace the corresponding region of SHIV-vpu⁺ env. The modified 3’ half was ligated with the 5’ half of SHIV-vpu⁺ proviral DNA to form full-length SHIV-1157i (Fig. 1a) (172, 358).

Animals and animal care

An infant Indian rhesus macaque (Macaca mulatta), RPn-8, was inoculated intravenously with 6 ml cell-free supernatant from 293T cells transfected with the infectious molecular clone, SHIV-1157i, and followed prospectively with complete blood counts, T-cell subset analyses, viral RNA load determinations, and clinical exams (172). This animal was kept according to National Institutes of Health guidelines on the care and use of laboratory animals at the Yerkes National Primate Research Center (YNPRC). The facility is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International. All experiments were approved by the Animal Care and Use Committees of the YNPRC and the Dana-Farber Cancer Institute.

Plasma viral RNA load

RNA was isolated from plasma using QiaAmp Viral Mini Kit (Qiagen, Valencia, California, USA), and vRNA loads were measured by quantitative reverse transcriptase
PCR for SIV gag sequences (164). The detection limit was 50 viral RNA copies/ml plasma (172).

**PCR, cloning and sequencing**

Sample collection, cloning and sequencing of infant 1157i V1–V5 env region have been described previously (438, 439). Genomic DNA of infected macaque peripheral blood mononuclear cells (PBMC) was extracted from 6, 20, 37, 50 and 64 month postinoculation samples using Genomic DNA Purification Kit (Gentra Systems, Valencia, California, USA). Nested PCR was used to amplify a 1.1 kb fragment spanning the V1–V5 env region. First-round PCR was performed with the primers ENF1 (5’ GATGCATGAGGATATAATCAGTTTATGGGA 3’) and ENR1 (5’ ATTGATGCTGCGCCCATAGTGCT 3’). Second-round PCR was performed with the primers 1157i-DraIII (5’ TTGACTCCACTCTGTGTCACTTTAAAG 3’) and 1157i-AvrII-AS (5’ TGCTATTCTAGGGGCTTGATTTTCTAC 3’). To minimize PCR bias, first-round PCR products were generated in duplicate and combined to be used as templates for the second-round PCR. Amplified fragments were cloned into pSP72 NL4-3 AS-Av after digestion with DraIII and AvrII and sequenced with dideoxy terminators (ABI BigDye Kit).

**Sequence analysis**

Sequence alignment was carried out on the translated amino-acid sequence in ClustalW (378), as implemented in BioEdit 7.0.9.0. Neighbor-joining phylogenetic analyses were done in MEGA (Molecular Evolutionary Genetics Analysis) using the Kimura 2-parameter distance to explore genealogical relationships among infant 1157i
and macaque RPn-8 V1–V5 sequences, and support for the nodes was evaluated with bootstrap.

Variations in genetic diversity, genetic divergence, the number and location of putative N-linked glycosylation sites (PNGSs) and the length of the V1–V5 fragment were analyzed. Viral genetic diversity was estimated as the average nucleotide difference between sequences within a contemporaneous set, and genetic divergence was calculated as the average genetic distance to the earliest viral population collected for the human or macaque. The number and location of PNGSs were estimated using N-GlycoSite from the Los Alamos National Laboratory.

The instantaneous rates of nonsynonymous (dN) and synonymous substitutions (dS) were compared to evaluate the role of natural selection. Estimates of dN, dS, and dN/dS for each time point were obtained in Datamonkey using the Fixed Effects Likelihood procedure (204).
Results

HIV1157i and SHIV-1157i infection courses

The HIV-C V1–V5 region used to construct SHIV-1157i was derived from a cloned viral isolate of infant 1157i obtained at 6 months. This child was infected *in utero*, delivered naturally with normal birth weight and breast-fed till 20 months old. Both the child and his mother were antiretroviral drug naive. Infant 1157i is designated as a slow progressor as he remained clinically asymptomatic throughout the study period, with a CD4⁺ T-cell count of 811 cells/ul at approximately 8 years. Analyses of viral isolates from this infant have been published previously (438), and the viral sequences are included here for comparison.

SHIV-1157i was constructed as described (Fig. 1a) (358) and inoculated intravenously into the infant macaque RPn-8. Systemic infection was achieved after inoculation with a plasma viral RNA load of approximately $3.7 \times 10^4$ copies/ml detected at 2 weeks postinoculation and remained at approximately $1.0 \times 10^4$ copies/ml or higher for most of the study period (Fig. 1b). CD4⁺ T cells steadily declined over time, reaching less than 200 cells/ul at approximately 28 months postinoculation (Fig. 1c) and remained low till euthanasia at 64 months postinoculation because of AIDS, as characterized by opportunistic infections (172). PBMCs were collected longitudinally throughout the study, and samples from 6, 20, 37, 50 and 64 months postinoculation were selected for analysis. The V1–V5 region was cloned and sequenced. A total of 161 V1–V5 *env* clones were obtained from all time points, approximately 32 clones per time point.

Phylogenetic relationship among V1–V5 sequences from macaque and human
The neighbor-joining phylogeny clustered all envelope sequences derived from the infected macaque in a monophyletic clade (Fig. 2), consistent with the fact that they all derived from a single infectious molecular clone, SHIV-1157i. Both the macaque and infant portions of the tree resemble the idealized trees expected for viruses evolving under continuous immune pressure (148). Sequences collected at later time points concentrated on the longer branches of the tree. The distribution of sequences from later time points indicates that the phylogenetic structure of the tree is more complicated in the infant than in the macaque. Infant sequences from the 67-month time point are scattered throughout the tree, with some sequences clustered with earlier lineages. In contrast, all macaque sequences from the last time point were found in a monophyletic clade.

**Longitudinal variation in diversity and divergence**

Constraints within the viral $env$ may hamper certain amino acid changes, regardless of any unique selective pressure by the individual host species, because of the cost to replicative fitness of the virus. As such, $env$ changes we observed in infant 1157i during viral evolution may be similarly reflected in the macaque. To determine whether longitudinal variations of the V1–V5 region in isolates from RPin-8 and 1157i have similarities, we compared changes in the diversity and divergence of all viral populations sampled from each host. Viral genetic diversity measures the level of polymorphism found within a viral population at the nucleotide level, whereas divergence measures the differentiation of a population relative to the original strain. In the case of macaque, divergence was calculated relative to the sequence of the single clone inoculated experimentally, whereas in the infant, divergence was calculated as the average genetic
distance between all sequences within the viral population at a given time point and all sequences from the viral population at birth.

Our results indicate that diversity and divergence followed similar trajectories in both the macaque and infant at early time points but differ at later time points (Fig. 3a and b). In the case of the infant, diversity increased quasimonotonically over time, reaching an average pair-wise distance of 4.3% at 67 months after birth. Diversity in the macaque increased in parallel with the infant for the first 20 months. Thereafter, diversity gradually leveled off, reached its peak of an average genetic distance of 2.8% in the population collected 50 months postinoculation and decreased slightly at the time of euthanasia (Fig. 3a).

Sequence divergence in the infant increased in parallel with its diversity until it reached its maximum of 3.6% at 48 months. This was ensued by a noticeable decrease to 2.7% at 67 months. This decrease in divergence coincided with the reemergence of viral lineages derived from earlier sequences shown by our phylogenetic results. Interestingly, divergence in the macaque increased monotonically throughout the study and reached 3.7% by 64 months postinoculation (Fig. 3b).

**Synonymous and nonsynonymous changes in macaque and human**

To evaluate the role played by natural selection, we compared the ratio of dN/dS (Fig. 3c). Briefly, dN/dS ratios are expected to be approximately 0 under strong purifying selection, approximately 1 under neutrality and more than 1 in cases in which positive Darwinian selection is the major evolutionary force driving variation in the gene studied. In the infant, the dN/dS ratio was more than 1 in the majority of the populations sampled,
ranging from 0.97 at 6 months to 1.44 at 67 months of age. In contrast, the dN/dS ratio in the macaque remained less than 1 till 50 months postinoculation during which it declined to 0.52. These data suggest that selective pressure was stronger in the infant than in the macaque. We then compared the distribution of these changes in the viral populations from the infant and macaque by plotting the number of observed dN per codon per time point (Fig. 3d). The comparison shows that despite the differences in the number of changes and selective pattern, amino-acid replacements accumulated in similar regions in the infant and macaque. These regions include the variable loops as well as the supposedly conserved C3 and C4 regions.

**Longitudinal changes in the number of N-linked glycosylation sites in V1–V5**

It has been suggested that lengthening of the V1–V5 region and an increase in PNGSs correlate with slower disease progression (31, 322). To see if this held true in our study, we examined the change in number of PNGSs in RPn-8 and found that it declined over time from 26 PNGSs present in the inoculum to a mean of 21 PNGSs by 37 months postinoculation (Table 1). This was followed by a gradual recovery to a mean of 25 PNGSs by 64 months postinoculation. The V1–V5 length in RPn-8 followed a similar trend (Table 1). In contrast to the macaque, variation in PNGSs and V1–V5 length in the infant showed no directional pattern.

In both the macaque and the infant, the majority of the PNGSs were distributed in identical regions, with 17 PNGSs in RPn-8 and 14 PNGSs in 1157i that were relatively conserved throughout the infection (Fig. 4a). These conserved PNGSs may have important roles in the structural integrity or receptor binding capability of envelope.
However, differences in PNGS do exist between the macaque and infant. The macaque had only nine PNGSs that were variable throughout the infection, compared with 12 PNGSs in the infant. Interestingly, six of these variable sites (N56, 100, 165, 208, 324 and 327) were identical between the macaque and infant.

The temporal dynamics of the variations in PNGSs was similar for both the infant and macaque. For instance, sites N100, 208, 324 and 327 fluctuated in both 1157i and RPn-8 over time. For site N56, changes in the infant were observed starting from 6 months. In comparison with the macaque, the identical PNGS was ablated at 37 months but gradually reemerged by 64 months postinoculation. Another site with similar changes between 1157i and RPn-8 is N165. Prevalence of the PNGS at N165 reached its lowest level by 36 months in the infant before remerging. Likewise, this site was reduced in the macaque at 20 months, but in contrast with the infant, this PNGS was completely ablated by 37 months postinoculation and never reemerged at later time points in the macaque.

We also noticed a pattern unique to the macaque, in which the prevalence of several PNGSs, including N56, was reduced near 37-month postinoculation but gradually reemerged at later time points.
Discussion

In the present study, we compared changes in the V1–V5 envelope region in viral populations sampled longitudinally from an R5 SHIV-C-infected macaque with those from a Zambian infant. Longitudinal variation in sequence diversity and divergence in HIV-1 infection in humans has been intensely studied, and we found that SHIV infection in this macaque followed the general pattern described for HIV-1 infection in humans (347): an initial phase in which both diversity and divergence increased linearly up to the 20-month sampling interval; an intermediate phase in which diversity stabilized but divergence continued to increase up to the 50-month sampling interval; a late phase in which divergence and diversity decreased or stabilized from the 50-month sample onwards. Our sampling intervals did not allow us to precisely determine the time transitions between the different phases.

In humans naturally infected with HIV-1, differential rates of disease progression have been linked with viral genetic diversity and dN/dS ratios. In particular, higher dN/dS estimates appear to correlate with slower rates of disease progression, as is the case for having a higher number of sites inferred to be evolving under positive Darwinian selection or having high adaptation rates (317, 414). Estimates of genetic diversity and dN/dS were all higher in the infant, a slow progressor, which is in agreement with previous observations (250, 327, 385). These data suggest that ablation of the macaque’s immune responses at later time points resulted in a drastic reduction of selective pressure on the viral populations. In comparison, the infant’s immune status appears to be healthy and exert a significant selective pressure on the viral populations; in fact, the child remains asymptomatic at age 10 years.
The distribution of dN along the V1–V5 region for the infant and macaque was located primarily within the variable loops. But our analysis also indicated an unexpected variable domain located in the α2-helix of C3 in the infant and macaque viruses. Alterations within this domain correlate with low CD4 T-cell counts in the macaque, suggesting that it may contain an immune epitope and could affect the biological properties of the virus (264, 267, 316, 329, 335). This phenomenon could be clade specific, as the homologous region is reported to be more conserved in clade B (71, 130, 138, 264). As reported by other groups, emergence of deletions within V4 was also observed in the infant and macaque in this study (Fig. 4b) (24, 42, 165).

N-Glycosylation plays a critical role in immune evasion, and increases in the number of PNGSs have been linked with immune resistance (70, 93, 133, 405). Although we found no clear pattern of change for the number of PNGSs in the infant, we did observe an interesting pattern of decreasing PNGSs after inoculation and a gradual recovery at later time points in the macaque. Reductions of PNGS number in the macaque between 6 and 37 months postinoculation coincide with its rapid decline in CD4 T-cell count and the onset of AIDS. Whether the eventual recovery of PNGSs at the last time point is because of immune escape or increase in replication fitness is unclear (151, 278, 412, 421).

Besides sharing similar PNGS distribution patterns and hot spots, we observed several similar PNGS variations within viral populations of the infant and macaque. In particular, N56 that locates at the C-terminal of V2 and N165 that locates within C2. Alteration of PNGS at the C-terminal of V2 has been reported in animals infected by clade B SHIV (31, 165, 239). A recent study also demonstrated that removal of PNGS at
this position increases HIV-1 89.6 sensitivity to neutralizing antibodies (230), thus suggesting that N56 could be a common immunological epitope for both clade B and C viruses (118, 369). It is surprising to find a variable PNGS, N165, within a constant region in the infant and macaque. Changes at this site coincide with the macaque’s persistently low CD4 T-cell counts and were shown by others to affect the virus sensitivity to monoclonal antibodies (254, 358). Ablation of a similar site has also been reported in patients (405). It is possible that N165 is an immune epitope, and ablation of it might enhance the infectivity or kinetics of the escape viruses.

In order to eliminate the possibility that any of the clones were the result of \textit{in-vitro} PCR recombination, phylogenetic analyses were performed to distinguish between PCR recombination and actual \textit{in-vivo} recombination events. Briefly, with \textit{in-vivo} recombination, one would observe an accumulation of mutations after a recombination event on a given branch within an estimated phylogeny. However, with PCR recombination, there would be no such additional substitutions on the recombinant branch, and therefore the clones would be nonunique and identical to parental clones. Application of this test revealed that a negligible number of clones were indeed generated by PCR recombination. In fact, only two and four clones were identified for the macaque and infant, respectively, that showed no additional substitutions along a branch in the estimated phylogeny (data not shown). Thus, PCR recombination occurred on a minor scale and did not adversely affect our analyses. Limiting dilution PCR (LDPCR) was not feasible in this study due to insufficient materials for most of the time points analyzed. Nevertheless, LDPCR was carried out on the last time point sampled of the macaque and revealed clones with similar phylogenetic distributions that were consistent with those
obtained by the regular PCR approach. No inherent differences were observed between the two procedures.

Taken together, our data suggest that genetic evolution of an R5 SHIV generated from the envelope of a primary isolate reflects changes that occur in HIV-1 during disease progression in the infected infant. The infected macaque has a relatively faster disease progression in comparison with the infant but a much slower disease progression compared with other SHIVs such as SHIV89.6P. This could be due to the differences in selection forces between the two species. Our data also imply that these changes occur in a compressed time frame since the infected macaque progressed to AIDS while the child has remained asymptomatic. Although there is no identical individual amino-acid mutation comparable with previously reported clade B studies, mutations do seem to occur in similar regions. In addition, even though it is difficult to make a direct comparison of the time course between human and macaque in disease progression, mutations observed in our infected macaque at later time points could be indicators of future disease progression in the child. A further understanding of factors that caused these genetic changes could provide beneficial insights for future HIV-1 vaccine designs, and the significance of these changes on the biological function of the envelope is currently under investigation. The caveat for our study is the small sample size, and the observation may be limited to this human and macaque pair. Nevertheless, this is a prospective study and the first study to our knowledge demonstrating the evolutionary similarity of a primary clade C envelope between a naturally infected human with an experimentally infected macaque. This study also reinforces the previous observation that
HIV-1 envelope undergoes similar changes in human and monkey during disease progression (165).
Acknowledgements

This work was supported by PHS grants, CA75903, TW001429 and NCRR COBRE grant RR15635 to C.W.; P01AI48240 to R.M.R., C.W. and R.A.R.; R01 DE12937, R01 DE0160354 and R37AI34266 to R.M.R.; P.L. was supported by an FWO postdoctoral research fund.

F.Y.T performed experiments, analyzed data and wrote the manuscript. F.G.H, D.C.T., P.L. and H.Z. helped analyze data. R.A.R oversaw the animal aspects of the study. R.M.R. and C.W. oversaw all aspects of this project. All authors provided critical input on the manuscript.
Figures and Figure Legends

Figure 1.

(a) SHIV-vpu+
Figure 1. (a) Pvu I (P) was introduced into the 3’ half of SHIV-vpu⁺ (172) proviral DNA. The 2.0 kb KpnI-PvuI fragment of HIV1157i [spanning most of gp120 as well as the entire gp41 extracellular domain and the transmembrane region (TM)] was amplified to replace the corresponding region in the SHIV-vpu⁺ envelope. The modified 3’ half of SHIV-vpu⁺ was ligated with the 5’ half of SHIV-vpu⁺ proviral DNA to form the full-length SHIV-1157i (172, 358). (b) Plasma viral RNA load and (c) absolute CD4⁺ T-cell counts in animal RPN-8 after viral inoculation.
Figure 2.
Figure 2. Consensus tree from a neighbor-joining bootstrap analysis showing phylogenetic relationships among viral samples derived from the longitudinal follow-up of the macaque (RPn-8, solid lines) and infant (1157i, dashed lines) in this study. Labels indicate the source and time of sample collection. For example, RPn-8 20M corresponds to viral sequences coming from the macaque and collected 20 months after inoculation. Cut-off value for the condensed tree was set at 75%.
Figure 3.
Figure 3. Changes in (a) genetic diversity and (b) divergence over time for infant 1157i and macaque RPn-8. Genetic diversity is calculated from the average number of nucleotide differences within a given time point. Genetic divergence is calculated from the average number of changes between each time point and the initial population. (c) nonsynonymous and synonymous (dN/dS) ratio over time for infant 1157i and macaque RPn-8. (d) Estimated number of observed nonsynonymous substitutions per codon within V1–V5 region in infant 1157i and macaque RPn-8. Results represented are cumulative of all time points for human or macaque. M represents either months of age in the human infant or months after inoculation in the macaque. Numbers on the horizontal axis correspond to amino-acid position within the sequence alignment. All variable loops and constant regions within the alignment are labeled.
Figure 4. (a) V1–V5 consensus sequence of infant 1157i at 0 month of age, and inoculation strain SHIV-1157i is shown. PNGSs within sequences from all time points were located as described. n represents relatively conserved PNGSs in macaque or human over time. N represents variable PNGSs in macaque or human over time. Common variable PNGSs between macaque and human are circled. Variable regions and constant regions are shown along the bottom of the sequences. (b) Sequence alignment of V4 in infant 1157i and macaque RPn-8 at different time points. Sequences were aligned using Bioedit 7.0.9.0. Sequences represented here are examples from each time points. These are not consensus sequences and do not represent all observed variations within V4.
Table 1. Putative N-glycosylation site numbers and length of V1-V5 region.

<table>
<thead>
<tr>
<th>Sample</th>
<th>n</th>
<th>V1 length</th>
<th>V2 length</th>
<th>C2 length</th>
<th>V3 length</th>
<th>C3 length</th>
<th>V4 length</th>
<th>C4 length</th>
<th>V5 length</th>
<th>Total V1-V5 length</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHIV-1157I</td>
<td>1</td>
<td>22</td>
<td>42</td>
<td>99</td>
<td>35</td>
<td>52</td>
<td>28</td>
<td>40</td>
<td>15</td>
<td>333</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>n</th>
<th>V1 PNGS</th>
<th>V2 PNGS</th>
<th>C2 PNGS</th>
<th>V3 PNGS</th>
<th>C3 PNGS</th>
<th>V4 PNGS</th>
<th>C4 PNGS</th>
<th>V5 PNGS</th>
<th>Total V1–V5 PNGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>11571 00M</td>
<td>31</td>
<td>3 (3–3)</td>
<td>3 (2–3)</td>
<td>8 (7–8)</td>
<td>1 (1–1)</td>
<td>3 (2–3)</td>
<td>4 (3–4)</td>
<td>2 (1–2)</td>
<td>2 (2–2)</td>
<td>26 (25–26)</td>
</tr>
<tr>
<td>11571 06M</td>
<td>29</td>
<td>3 (2–4)</td>
<td>3 (1–3)</td>
<td>7 (7–8)</td>
<td>1 (0–2)</td>
<td>3 (3–3)</td>
<td>4 (3–4)</td>
<td>2 (1–2)</td>
<td>2 (1–2)</td>
<td>25 (23–27)</td>
</tr>
<tr>
<td>11571 29M</td>
<td>26</td>
<td>4 (3–6)</td>
<td>2 (1–3)</td>
<td>7 (6–8)</td>
<td>1 (1–1)</td>
<td>2 (2–3)</td>
<td>4 (3–4)</td>
<td>2 (1–2)</td>
<td>2 (2–2)</td>
<td>24 (21–27)</td>
</tr>
<tr>
<td>11571 36M</td>
<td>26</td>
<td>4 (3–6)</td>
<td>3 (2-3)</td>
<td>6 (5–7)</td>
<td>1 (1–1)</td>
<td>2 (2–3)</td>
<td>4 (3–4)</td>
<td>2 (1–2)</td>
<td>1 (1–2)</td>
<td>23 (20–25)</td>
</tr>
<tr>
<td>11571 48M</td>
<td>25</td>
<td>4 (3–5)</td>
<td>2 (1–3)</td>
<td>7 (6–8)</td>
<td>1 (1–1)</td>
<td>2 (1–3)</td>
<td>4 (4–4)</td>
<td>1 (0–2)</td>
<td>1 (0–2)</td>
<td>23 (20–25)</td>
</tr>
<tr>
<td>11571 67M</td>
<td>32</td>
<td>4 (3–7)</td>
<td>3 (2–3)</td>
<td>7 (6–8)</td>
<td>1 (1–1)</td>
<td>3 (1–3)</td>
<td>4 (3–4)</td>
<td>2 (1–2)</td>
<td>2 (1–2)</td>
<td>25 (21–28)</td>
</tr>
<tr>
<td>SHIV-1157I</td>
<td>1</td>
<td>3 (3)</td>
<td>3 (3)</td>
<td>3 (3)</td>
<td>4 (4)</td>
<td>2 (2)</td>
<td>2 (2)</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPh-8 06M</td>
<td>34</td>
<td>3 (2–3)</td>
<td>3 (2–3)</td>
<td>8 (7–8)</td>
<td>1 (1–1)</td>
<td>3 (3–3)</td>
<td>4 (3–4)</td>
<td>2 (1–2)</td>
<td>2 (2–2)</td>
<td>25 (24–26)</td>
</tr>
<tr>
<td>RPh-8 20M</td>
<td>31</td>
<td>3 (2–3)</td>
<td>3 (3–3)</td>
<td>7 (6–8)</td>
<td>1 (1–1)</td>
<td>3 (2–3)</td>
<td>4 (4–4)</td>
<td>2 (1–2)</td>
<td>1 (0–1)</td>
<td>24 (23–25)</td>
</tr>
<tr>
<td>RPh-8 37M</td>
<td>30</td>
<td>3 (2–3)</td>
<td>2 (2–2)</td>
<td>6 (5–7)</td>
<td>1 (0–1)</td>
<td>3 (3–3)</td>
<td>4 (3–4)</td>
<td>2 (2–2)</td>
<td>1 (1–1)</td>
<td>21 (19–23)</td>
</tr>
<tr>
<td>RPh-8 50M</td>
<td>30</td>
<td>2 (2–3)</td>
<td>2 (1–2)</td>
<td>7 (6–7)</td>
<td>1 (0–1)</td>
<td>3 (1–3)</td>
<td>4 (4–4)</td>
<td>2 (2–2)</td>
<td>1 (0–2)</td>
<td>22 (20–24)</td>
</tr>
<tr>
<td>RPh-8 64M</td>
<td>36</td>
<td>4 (3–4)</td>
<td>2 (2–3)</td>
<td>7 (6–7)</td>
<td>1 (1–1)</td>
<td>3 (2-3)</td>
<td>4 (3–5)</td>
<td>2 (2–2)</td>
<td>2 (1–2)</td>
<td>25 (23–27)</td>
</tr>
</tbody>
</table>

Number of samples (n). Numbers reported are to the nearest whole number. Numbers are given as mean (minimum–maximum). PNGS, putative N-linked glycosylation site.
Chapter 3

Variations in the biological functions of HIV-1 clade C envelope in a SHIV-infected rhesus macaque during disease progression

For Yue Tso\textsuperscript{a}, Levon Abrahamyan\textsuperscript{a}, Shiu-Lok Hu\textsuperscript{b,c}, Ruth M. Ruprecht\textsuperscript{d,e} and Charles Wood\textsuperscript{a*}

\textsuperscript{a}Nebraska Center for Virology and the School of Biological Sciences, University of Nebraska-Lincoln, Nebraska 68583; \textsuperscript{b}Department of Pharmaceutics and \textsuperscript{c}The Washington National Primate Research Center, University of Washington, Seattle, Washington 98121; \textsuperscript{d}Dana-Farber Cancer Institute and \textsuperscript{e}Harvard Medical School, Boston, Massachusetts 02215

*Corresponding author

Mailing address: Nebraska Center for Virology, School of Biological Sciences, University of Nebraska-Lincoln, Morrison Center P.O. Box 830666, Lincoln, NE 68583-0900

Phone: (402) 472-4550; Fax: (402) 472-3323; E-mail: cwood@unlnotes.unl.edu

Data submitted for review
Abstract

Objectives: To examine how the biological functions of HIV-1 clade C (HIV-C) envelopes from a SHIV-infected rhesus macaque evolve with disease progression.

Design: The V1-V5 envelope regions were cloned from a rhesus macaque infected with SHIV-1157i, an infectious molecular clone that expresses a recently transmitted HIV-C envelope from Zambia. Envelope clones from various disease stages were examined for biological properties and their association with disease progression.

Methods: We assessed envelope charges by bioinformatics analysis, envelope neutralization sensitivity and replicative fitness by infection of PBMC with provirus expressing different V1-V5 clones. Cell surface-associated envelope and gp120 content on virion were quantified by cell-surface biotinylation of producer cells and gp120 ELISA of purified virions, respectively. We also assessed envelope fusion capacity and CD4 binding affinity.

Results: Envelopes from early infection tended to be more positively charged and neutralization sensitive, but the net charge decreased and Env became neutralization resistant with time. Early envelopes also exhibited better cleavage, higher gp120 content on virion, superior fusion capacity and faster replication kinetics compared with envelopes from later time points. Lastly, CD4 binding tended to improve as disease progressed, but reverted to early infection level after the onset of AIDS.
Conclusion: Envelope gained distinctive biological characteristics as disease progressed. Potentially beneficiary functions could have arisen during early infection for better *in-vivo* viral fitness. However, some Env properties reverted to pre-AIDS levels in the absence of immune pressure.
Introduction

Despite years of efforts, there is no efficacious vaccine against HIV-1. This impasse is partly due to the highly variable nature of the HIV-1 envelope gene, which enables viral escape from immune surveillance (326). Most of the envelope diversity is within the V1-V5 region, which encompasses multiple crucial functions, such as binding to receptor and co-receptors, and can influence envelope fusion to the cell membrane (108, 205, 383, 419). Being the major surface component of HIV-1 envelope, V1-V5 is also constantly targeted by host humoral responses (322, 405). Given these important roles, the V1-V5 region is frequently targeted for diverse anti-HIV-1 strategies.

Understanding the changes of HIV-1 envelopes in vivo will be of tremendous importance. The first step towards achieving this goal is to examine how the biological functions of HIV-1 V1-V5 region evolve during disease progression. Unfortunately, there is no consensus as to what these changes might be, due to the difficulties in following individuals from initial HIV-1 infection throughout disease progression.

To address this question, we have previously shown a strikingly similar molecular evolution of the HIV-1 clade C (HIV-C) envelope during infection in a Zambian infant (1157i) and an infant rhesus macaque (RPn-8) infected by SHIV-1157i, a SHIV expressing the recently transmitted HIV-C envelope of 1157i (387). This finding validated our macaque model for studying envelope evolution during disease progression. Importantly, our macaque model allows studying the longitudinal changes of the biological functions of the V1-V5 region from an infectious molecular clone inoculum over the entire disease course. Moreover, significant mutations were observed within the HIV-C envelope sequences from different disease stages in the infected macaque RPn-8
(387). However, whether these envelope mutations translate into different biological functions remains unanswered. Here, we addressed this question by characterizing the biological properties of the V1-V5 region cloned from RPn-8 during disease progression. Our hypothesis is that envelopes from various disease stages have different biological characteristics, which might confer fitness advantages for the virus.

This study will be of substantial importance since it is the first reported clade C SHIV (SHIV-C) model that documented progression to AIDS. Moreover, the recently transmitted Zambian HIV-C primary isolate we utilized is CCR5 tropic and represents the dominant clade of HIV-1 infection worldwide (69, 92, 158, 298). The observed changes in biological functions in this setting should reflect those occurring during natural disease progression.
Materials and methods

Animal and animal care

Details of animal care and procedures were described (172, 387). Briefly, an infant Indian rhesus macaque (RPn-8) was intravenously inoculated with SHIV-1157i, a SHIV-C expressing the envelope of a recently transmitted Zambian pediatric HIV-C. The infected animal was monitored from inoculation until euthanasia due to AIDS (~5 years). The animal was housed and cared for according to the National Institutes of Health Guidelines on the Care and Use of Laboratory Animals at the Yerkes National Primate Research Center (YNPRC), which is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International. All animal procedures were approved by the Animal Care and Use Committees of the YNPRC and the Dana-Farber Cancer Institute.

Construction and preparation of viruses

The SHIV-1157i inoculum was prepared as described (172, 358). Genomic DNA was extracted from PBMC samples of RPn-8 collected at 6, 20, 37, 50 and 64 months post-infection (mpi). The V1-V5 region was amplified from genomic DNA and cloned into the envelope expression (pSRH NLA/S/Av) and provirus (pNL4-3 A/S/Av) plasmids as described (387, 440, 441). Thus, both types of plasmids bear a chimeric HIV-1 NL4-3 envelope with its V1-V5 region derived from RPn-8. Infectious virus was generated by transfection of 293T cells with the proviral plasmids, supernatant was collected and filtered at 48 hrs post-transfection (pt). For purified virus, the filtered supernatant was
concentrated by ultracentrifuge and the resulting virus pellet was re-suspended in their respective assay medium.

**Envelope Charge analysis**

V1-V5 region sequences were obtained from the previous study (387). Charged residues of the envelope were computed using AminoTrack™ (241). The value of +1 was assigned to positively charged arginine and lysine. The value of -1 was assigned to negatively charged aspartic acid and glutamic acid. GraphPad Prism 5 (GraphPad Software) was used for statistical analysis.

**Neutralization assay**

Pseudotyped viruses were generated by co-transfection of COS-1 cells with the envelope expression and backbone pNL4-3-deltaE-GFP plasmids, supernatants were collected and filtered at 48 hrs pt. Pseudotyped virus encoding NL4-3 and SV-A-MLV envelopes served as controls. Next, 1x10^4 TZM-bl cells/well were seeded into 96-well plates 24 hrs before the assay. Heat-inactivated plasma samples of RPn-8 were 3-fold serially diluted starting from 1:20. Pooled monoclonal antibodies (mAbs) were prepared by mixing equal concentrations of the mAbs 2F5, 4E10, 2G12 and b12, followed by 3-fold serially dilutions starting from 80 ug/ml (total concentration of all mAbs in the mixture). An inoculum of 300 50% tissue culture infectious doses (TCID_{50})/ml of pseudotyped viruses was supplemented with 80 ug/ml DEAE-dextran and mixed with an equal volume of either the serially diluted heat-inactivated plasma or pooled mAbs. The mixture was incubated for 1 hr at 37 °C before addition to TZM-bl cells in triplicates. After 48 hrs incubation at 37 °C, the cells were washed, lysed with Luciferase Assay
System reagents (Promega) and its luciferase activity measured by luminometer. The 50% inhibitory concentration (IC\textsubscript{50}) in ug/ml and reciprocal IC\textsubscript{50} were calculated using GraphPad Prism 5.

**Viral replication kinetics**

Infectious viruses bearing the V1-V5 region of RPn-8 were used to infect PBMC from HIV-1-seronegative donors. PHA-stimulated PBMC (3.5x10\textsuperscript{5} cells) were infected at a multiplicity of infection (MOI) of 0.05 overnight at 37 °C in duplicates. The next day, the infected cells were washed and resuspended in RPMI growth medium containing human interleukin-2 (hIL-2) (Roche Applied Science). Cell-free supernatants were collected every 3 days. The level of reverse transcriptase (RT) activity in 10 µl of collected supernatants was determined in triplicates by the standard RT activity assay and expressed in counts-per-minute (CPM) (162). The experiment was repeated using a different HIV-1-seronegative donor PBMC.

**Quantification of virus-associated gp120**

The amount of virus-associated HIV-1 surface glycoprotein was measured using ELISA as described (29). Briefly, 96-well plates were coated overnight with anti-gp120 antibody D7324 (Aalto Bio Reagents Ltd), followed by washing with Tris-buffered saline (TBS) and blocked with nonfat-milk solution. Purified virus (100 ng/ml p24 equivalents) was lysed and added to the plates in triplicates for 5 hrs incubation at 37 °C. Unbound proteins were removed by extensive washes. For detecting captured gp120, mAb T43 (NIH) was added at room temperature. Importantly, the mAb T43 epitope lies within the envelope C1 region, which is identical among all our samples. The plates were washed
and captured gp120 was detected by addition of goat anti-mouse horseradish peroxidase (Jackson ImmunoResearch Laboratories) with 1-Step Ultra-TMB-ELISA substrate (Pierce). Gp120 quantity was measured by microplate reader and calculated by comparison with a standard curve generated using serial dilutions of purified HIV-C 1157ip gp120 with a known concentration.

**Quantification of cell surface-associated envelope**

Labeling of cell-surface proteins with biotin was described (7). Briefly, 293T cells transfected with proviral plasmids were harvested at 48 hrs pt, washed and incubated with 1 mg/ml of EZ-Link Sulfo-NHS-LC-biotin (Thermo Scientific) for 30 min on ice, followed by extensive washes and addition of lysis buffer containing proteinase inhibitor cocktail. Cellular debris was removed and the total protein concentration in the cleared lysates was measured using a Pierce BCA protein assay kit (Pierce). Equal amounts of total protein were used for HIV-1 envelope immunoprecipitation with HIV-IG (NIH) overnight at 4 °C. The immune complexes were pulled down with Protein A/G Ultra-Link Resin (Thermo Scientific), washed and boiled in 2X SDS-PAGE sample buffers. Resin-free supernatant containing the immunoprecipitated proteins was resolved by Western blot. Biotinylated proteins were detected using streptavidin-conjugated-680 secondary antibody (Li-Cor Biosciences) and visualized by Odyssey infrared imager (Li-Cor Biosciences). Respective band intensity of the cell surface-associated HIV-1 envelope was analyzed using Odyssey application software version 3.0.

**In-vitro CD4 binding**
Envelope CD4 binding capability was determined by ELISA. Briefly, purified virus (10 ng/ml gp120 equivalents) was lysed and captured with D7324-coated plates in triplicates. Serial dilutions of CD4-IgG2 with known concentrations were added and incubated overnight at room temperature. Unbound CD4-IgG2 was removed by extensive washes. CD4-IgG2 bound to gp120 was detected by addition of goat-anti-human horseradish peroxidase (Santa Cruz Biotechnology) with 1-Step Ultra-TMB-ELISA substrate (Pierce) and measured by microplate reader. Half-maximal binding concentration (ng/ml) of CD4-IgG2 was calculated for each sample.

**FRET-based virus-to-cell fusion**

The fluorescent resonance energy transfer (FRET)-based virus-to-cell fusion assay was described (53, 54). Briefly, infectious viruses bearing the V1-V5 region of RPh-8 and the β-lactamase HIV-1 Vpr fusion protein (BlaM-Vpr) were generated by co-transfection of 293T cells with the provirus, pAdVantage (Promega) and pMM310 plasmids (NIH). Supernatants were collected at 48 hrs pt and purified as described earlier. The p24 concentration of purified viruses were measured with HIV-1 p24 ELISA (Perkin-Elmer). Then, 2.5x10^5 SupT1/CCR5 cells (256) were incubated with 50 ng of p24 of the purified viruses for 2 hrs at 37 °C in 96-well plates in triplicates. The infected cells were washed and resuspended in the CCF2-AM loading solution from GeneBLAzer Detection Kits (Invitrogen) for 1 hr at room temperature in the dark. The CCF2-AM-loaded cells were washed, resuspended in 10% FBS medium with probenecid and incubated overnight at room temperature in the dark. The next day, CCF2-AM-loaded cells were washed and fixed for flow cytometry. Once the BlaM-Vpr viruses fused with the target cell, β-lactamase will cleave the CCF2-AM dye and alter its fluorescence
emission spectrum from 520 nm (green) to 447 nm (blue). This change in fluorescence emission was detected by Cytek DxP10 (Cytek Development) and analyzed using FlowJo software v7.9 (TreeStar). GraphPad Prism 5 was used for statistical analysis.
Results

Charged residues of the HIV-C V1-V5 region

Charged residues play an important role in protein-protein interactions by affecting the physiological interactions between molecules, and might influence HIV-1 envelope-antibody interaction, receptor binding as well as disease progression (34, 266, 311, 345). Therefore, it was interesting to determine whether the charges of V1-V5 region varied with disease progression in monkey RPn-8.

As shown previously, RPn-8 maintained a high viral load (>1x10⁴ copies/ml) throughout the disease course (387) and the CD4⁺ T-cell counts gradually declined to <200 cells/ul from ~28 mpi onwards. The animal was euthanized at 64 mpi due to severe opportunistic infections (387). Using ~160 sequences collected previously across the entire disease course (387), we analyzed the charged residue distribution within the V1-V5 region. Our analysis showed a significant increase in the net charge of V1-V5 region during early infection (6 mpi) compared to the inoculum (Fig 1A). However, the net charge decreased substantially by 37 mpi when the CD4⁺ T-cell counts fell to <200 cells/ul. Surprisingly, as RPn-8 had advanced AIDS, the net charge of the V1-V5 region recovered to pre-AIDS levels.

This pattern of change was mainly caused by the V1-V3 region. For example, C2 charges reached the lowest levels by 37 mpi but rebounded later (Fig 1B). Other regions, such as V5, contributed towards the recovery of envelope charges during advanced AIDS (Fig 1C).

Neutralization characteristics
To determine the neutralization characteristics of the V1-V5 region, selected envelope clones were tested against naïve, contemporaneous and non-contemporaneous plasma from RPn-8. Envelope selection was based upon fusion capacity and position in the phylogenetic tree to maximize the representation of various branches. About five functional envelope clones per time point were selected for further analysis (387).

We found that neither the naïve nor the contemporaneous plasma could neutralize any of the envelopes tested (data not shown). However, non-contemporaneous plasma was able to neutralize envelopes from earlier time points, but this capability tended to decrease with disease progression (Fig 2A). This was further confirmed by the examination of 6 mpi envelopes against plasma samples from all later time points, where the 64 mpi plasma had the weakest neutralizing ability (Fig 2B).

Since plasma from RPn-8 might contain different neutralizing antibodies, depending on the time points, a reference panel that consisted of pooled mAbs was used to examine the relative neutralization sensitivity among the envelopes. Our data show that envelopes from 6 mpi were mostly sensitive to mAbs (Fig 2C). However, the majority of the envelopes displayed increased mAbs resistance by 37 mpi. Surprisingly, mAbs-sensitive envelopes re-emerged at 64 mpi.

**Replication kinetics**

To determine if envelopes derived from different disease stages could affect the viral replicative fitness and its relationship with neutralization sensitivity, we examined the replication kinetics of viruses expressing various envelope clones in PBMC. Nine envelope clones were selected from 6 (early infection), 37 (early AIDS) and 64
(advanced AIDS) mpi. Envelope selection was based on the mAbs neutralization results to maximize representation of different neutralization characteristics. As shown in Table 1, viruses with 6 month envelopes (clones #26 and #37) displayed distinctively faster ex-vivo replication (P = 0.0008) than those with 64 mpi envelopes, averaging ~7 days to reach half-maximal viral replication using PBMC from two human donors. In contrast, viruses bearing envelopes from later time points showed slightly slower ex-vivo replication compared to the inoculum over time.

**Quantity of envelope on cell surfaces and gp120 on virions**

To further examine the possible reasons for faster ex-vivo replication of the viruses with early envelopes, we first evaluated the incorporation of envelopes into virions. By measuring the gp120 content on virions using ELISA, we showed that the quantity of surface glycoprotein on virions tended to decline with disease progression. Viruses with 64 mpi envelopes only contained an average ~25% (P < 0.0001) of gp120 content compared with the inoculum SHIV-1157i (Fig 3A).

This decreased level of virion-associated gp120 could be caused by differences in the expression or cleavage of envelope. To eliminate these possibilities, 293T cell surfaces were labeled with biotin, and the quantity of cell surface-associated gp160 and 120 was measured by Western blot. Our data indicated no significant differences in gp160 expression levels between all envelopes (Fig 3B); thereby excluding differences in gp160 expression on cell surfaces as a plausible cause. However, we did observe a decrease in cell surface-associated gp120 levels, which positively correlated with the decline of gp120 content on virions over time (Spearman r = 0.7295, P = 0.0202) (Fig
3C). Therefore, the decreased recruitment of surface glycoprotein onto virions may have been a consequence of having less cell surface-associated gp120, which subsequently caused an increase of the cell surface gp160 to 120 ratio over time (Fig 3D). This was further supported by the negative correlation (Spearman $r = -0.8909$, $P = 0.0011$) between the cell surface gp160 to 120 ratio and gp120 quantity on virions (Fig 3D). Together, our data strongly suggest a deficiency in cleavage or trafficking of cleaved envelope to cell surfaces with disease progression.

**Envelope binding to CD4**

Another possible factor that can influence ex-vivo replication is divergence in the envelope-CD4 binding capacity. To this end, our *in-vitro* CD4 binding assay showed that the CD4 binding proficiency of envelope improved with disease progression. At 37 mpi, the envelopes had significantly enhanced binding to CD4 compared with the inoculum (Fig 4A). Surprisingly, this enhanced CD4 binding capability was not maintained and reverted to pre-AIDS levels during advanced AIDS.

**Envelope fusion to target cell**

Fusion of viral envelope to target cell is a critical rate-limiting step that could affect ex-vivo replication. A FRET-based virus-to-cell fusion assay was used to examine the fusion efficiency of envelopes from various time points. Our result showed that viruses with 6 mpi envelopes, except #5, had significantly enhanced fusion abilities compared with the inoculum ($P = < 0.0001$) (Fig 4B). Envelopes isolated from 37 mpi displayed a diverse fusion capability, with 37 month #1 having fusion ability close to the 6 mpi and the rest having lower fusion ability compared to the inoculum (Fig 4B).
Additionally, this weak fusion phenotype dominated among envelopes isolates from advanced AIDS (Fig 4B). This suggests that envelope fusion ability increased during early infection, but gradually declined with disease progression.
Discussion

This study showed that the V1-V5 regions of HIV-C envelope underwent significant changes in their biological characteristics throughout disease progression. Envelopes from early infection were more positively charged, neutralization sensitive and had faster ex-vivo replication and enhanced fusion capability. During early AIDS, the envelopes became more negatively charged with diminished functional capabilities ranging from ex-vivo replication to incorporation into virions. However, these envelopes did have significantly improved CD4 binding ability and were more neutralization resistant. Interestingly, the envelopes showed multiple reversions in their biological properties during advanced AIDS. For example, their charge, neutralization sensitivity and CD4 binding efficiency had all reverted to pre-AIDS levels. However, other functions remained relatively poor.

Our results are surprising given the current belief that viral fitness increases over time due to envelope evolution (310, 385). It is possible that the previous studies were unable to capture the early infection in adult HIV-1 patients and used PBMC co-cultured viral quasi-species for biological analysis. Our study represents envelope isolates throughout the disease spectrum, focusing on V1-V5 region changes. At the early infection of 6 mpi, monkey infant RPn-8 had an unambiguously higher CD4\(^+\) T-cell counts of >2000 cells/ul, which was age-appropriate, and high viral loads compared to reports of other investigators, reflecting acute infection (387). Moreover, having an isogenic backbone allowed us to focus on the V1-V5 region without the influences from other viral genes. Our use of PBMC for ex-vivo replication also better represents the in-vivo scenario than the cell lines used in previous study (108). Alternatively, the different
maturation state of the immune system between adults and infants during the initial infection might affect the evolutionary pressure placed on the virus.

Our study suggests that during early infection, the host immune system is intact and exerts selective pressures on the viruses, as the relatively early plasma can easily neutralized the early envelopes. To counter this host constraint, the envelope evolved compensatory mutations that resulted in more efficient fusion and faster replication kinetics, increasing the likelihood for the virus to infect new cells before being neutralized by the immune system. This possibility that early envelopes might have better fitness was mathematically predicted previously and observed recently in HIV-1 patients (13, 416). Although the exact mechanism of this outcome is still controversial, the “Red Queen” hypothesis was proposed as a possible explanation, where beneficiary mutations evolved in responses to immune selection resulting in increased fitness of the early envelopes (13, 74, 396).

During early AIDS, some immune pressure persisted as evident by the presence of neutralization-resistant envelopes and was also reported by others (34, 45). Importantly, the appearance of neutralization-resistant phenotypes coincided with a significant drop in the net charges of V1-V5 region. This strongly suggests that variations in net charges could be an escape mechanism employed by the envelope to evade humoral responses beside its glycan shield (405). The high CD4 binding ability of the envelopes at this stage could be another compensatory mutation for more efficient infection in an environment with limited availability of CD4+ T cells.
Another unexpected finding from our study is that viral fitness linked to the V1-V5 region, as defined by the *ex-vivo* replication of isogenic viruses differing only in their V1-V5 regions, did not improve despite reversion of several envelope properties during advanced AIDS. Importantly, the gp120 content on cell surface and on virion gradually decreased as disease progressed. Additionally, with the exception of the CD4 binding site mutant 6 month #5, decreased levels of gp120 on virion correlated with the decline in fusion ability from early infection to advanced AIDS (Spearman $r = 0.7667$, $P = 0.0214$) (graph not shown). These changes could be evasive responses toward immune surveillance and ultimately affected the neutralization sensitivity as suggested earlier (32). The severe depletion of CD4$^+$ T cells and absence of immune selection might facilitate the accumulation of deleterious mutations, possibly yielding envelopes with lower fitness reminiscent to the effect of “Muller’s ratchet” hypothesis (59, 102, 271, 433).

Despite our analysis of only one animal longitudinally, this is the first study of a SHIV-C-infected macaque that progressed to AIDS. To our knowledge, this is also the first study that demonstrated a decrease of gp120 content on virion with disease progression. Although gp120 shedding might reduce the gp120 content on virion, it was unlikely in our case as suggested by other investigators (66). By focusing on the V1-V5 region, we cannot eliminate the role of other viral genes or envelope regions, such as the signal peptide and gp41, in the overall viral fitness (17). However, our data strongly suggest that V1-V5 region plays a major role in disease progression (214, 238). Also, the presence of mixed population at any time points due to reactivation of latent viral reservoirs cannot be excluded.
Our findings underscore the dynamic interaction between host immune selection and HIV-1 envelope evolution. Our emphasis on the V1-V5 region had shed light on the evolution of its unique properties during disease progression. Moreover, variations in the envelope CD4 binding and fusion ability might suggest some corresponding structural changes with disease progression, which could have substantial implications for the development of vaccines and small molecule inhibitors.
Acknowledgments

This work was supported by Public Health Service grants, CA75903, TW001429 and the National Institute of Medical Sciences, Center for Biomedical Research Excellence grant P30 GM103509 to C.W; R01 AI076170 to S.L.H; P01 AI48240 to R.M.R., S.L.H and C.W.; American Recovery and Reinvestment Act funding 3P01 AI048240-08S1 to R.M.R., C.W.

Authors’ contributions: F.Y.T and L.A. performed experiments and analyzed data. F.Y.T wrote the manuscript. S.L.H. provided reagent and discussions. R.M.R. and C.W. oversaw all aspects of this project. All authors provided critical input on the manuscript.

The following reagent was obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: pMM310 from Dr. Michael Miller (53, 381); pNL4-3-deltaE-EGFP from Drs. Haili Zhang, Yan Zhou, and Robert Siliciano (442); TZM-bl cells from Drs. John C. Kappes, Xiaoyun Wu and Tranzyme Inc (295); SV-A-MLV-env from Drs. Nathaniel Landau and Dan Littman (221); HIV-1 gp41 mAbs (2F5, 4E10) and gp120 mAb (2G12) from Dr. Hermann Katinger (302, 363, 384); HIV-1 gp120 mAb (IgG1 b12) from Drs. Dennis Burton and Carlos Barbas (313); HIV-IG from NABI and NHLBI. We would like to thank Dr. James A. Hoxie for SupT1/CCR5 cells (256), Drs. Earl Patricia and Bernie Moss for mAb T43. We thank Dr Marielle Cavrois for her advice with the FRET-based virus to cell fusion assay and Dr Francois Villinger for his assistance in manuscript preparation.

Conflicts of interest

There are no conflicts of interest.
Figures and Figure Legends

Figure 1.
Figure 1. Distribution of charged residues within the V1-V5 envelope region. Each symbol represents one V1-V5 sequence. SHIV-1157i is the inoculum envelope, followed by envelopes from 6, 20, 37, 50 and 64 months pi. (A) Net charge of the V1-V5 region. (B) Charge distribution within the C2 and V5 region respectively.
Figure 2.
Figure 2. Envelope neutralization characteristics. (A) Pseudotyped viruses with envelopes from 20 months pi were tested against non-contemporaneous plasma from 37 months pi, 37 months pi envelope clones against 50 months pi plasma and 50 months pi envelope clones against 64 months pi plasma. Each symbol represents a pseudotyped virus encoding an individual envelope clone from that particular time point. (B) Pseudotyped viruses with envelopes from 6 month pi were tested against plasma of 20, 37, 50 and 64 months pi. (C) Pseudotyped viruses with envelopes from various time points were tested against pooled mAbs. Envelopes from the infectious molecular clone inoculum (SHIV-1157i), murine leukemia virus (MLV) and NL4-3 served as controls. Each symbol represents a pseudotyped virus encoding an individual envelope clone from that particular time point.
Figure 3.

A. gp120 content on virion relative to inoculum (%)

B. Band intensity of gp160 on cell surface relative to inoculum (%)

C. Spearman $r = 0.7305$
P = 0.0202

D. Spearman $r = -0.8909$
P = 0.0011
Figure 3. Quantity of envelope on cell surfaces and gp120 on virions. (A) Quantities of gp120 on virions expressing selected V1-V5 region from each time point. (B) Relative levels of 293T cell surface-associated gp160. (C) Correlation between the levels of virion-associated gp120 (black bars) and 293T cell surface-associated gp120 (red dots). (D) Correlation between the levels of virion-associated gp120 (black bars) and the ratio of 293T cell surface-associated gp160-120 (red dots).
Figure 4.

A. Half-maximum for CD4-IgG2 binding (ng/ml)

B. Mock 6 month #37

- Unlabeled CCF2 (520 nm)
- Cleaved CCF2 (447 nm)

Fusion positive: 0.300% (58.0%)

Fusion positive cells normalized to inoculum (%)
Figure 4. Envelope binding to CD4 and virus fusion with target cells. (A) Purified viruses, expressing selected V1-V5 regions from each time point, were tested with in-vitro CD4 binding assay to determine their half maximum binding to CD4-IgG2. Each symbol for individual envelope clones represents an independent experiment. “*” represent statistical significant compare to the inoculum, SHIV-1157i. The value of 6 month #5 may not reflect the actual data since it falls beyond the maximum range of our assay. (B) Envelope fusion with target cells was measured using FRET-based virus-to-cell fusion assay. An example for the detection of fluoresce emission shift from green (fusion negative) to blue (fusion positive) by flow cytometry was shown above the graph. The graph shows percentage of fusion-positive cells for individual envelope clones from each time point relative to the inoculum, SHIV-1157i. All differences were statistically significant compared to the inoculum.
Table 1: Ex-vivo replication of infectious viruses, expressing varies V1-V5 region, in PBMC.

<table>
<thead>
<tr>
<th>Envelopes</th>
<th>Half-time of maximal viral replication (days)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Donor 1</td>
<td>Donor 2</td>
<td></td>
</tr>
<tr>
<td>SHIV-1157i</td>
<td>9.739</td>
<td>8.951</td>
<td></td>
</tr>
<tr>
<td>6 month #5</td>
<td>10.032</td>
<td>10.798</td>
<td></td>
</tr>
<tr>
<td>* 6 month #26</td>
<td>7.099</td>
<td>7.083</td>
<td></td>
</tr>
<tr>
<td>* 6 month #37</td>
<td>6.510</td>
<td>6.315</td>
<td></td>
</tr>
<tr>
<td>37 month #1</td>
<td>10.555</td>
<td>10.813</td>
<td></td>
</tr>
<tr>
<td>37 month #11</td>
<td>8.998</td>
<td>9.797</td>
<td></td>
</tr>
<tr>
<td>37 month #S1</td>
<td>11.207</td>
<td>12.258</td>
<td></td>
</tr>
<tr>
<td>64 month #13</td>
<td>11.922</td>
<td>15.105</td>
<td></td>
</tr>
<tr>
<td>64 month #30</td>
<td>9.880</td>
<td>11.130</td>
<td></td>
</tr>
<tr>
<td>64 month #46</td>
<td>10.414</td>
<td>12.496</td>
<td></td>
</tr>
</tbody>
</table>

*Denote statistical significances (P = 0.0008) in comparison with the 64 month envelopes.
Chapter 4

Dynamics of Envelope Evolution in Clade C SHIV-infected Pig-tailed Macaques during Disease Progression Analyzed by Ultra-deep Pyrosequencing

For Yue Tso¹,‡, Damien C. Tully¹,†,‡, Sandra Gonzalez¹, Christopher Quince⁶, On Ho², Patricia Polacino³, Ruth M. Ruprecht⁴,⁵, Shiu-Lok Hu²,³ and Charles Wood¹*

¹Nebraska Center for Virology and the School of Biological Sciences, University of Nebraska-Lincoln, Nebraska, USA; ²Department of Pharmaceutics, ³The Washington National Primate Research Center, University of Washington, Seattle, Washington, USA; ⁴Dana-Farber Cancer Institute and ⁵Harvard Medical School, Boston, Massachusetts, USA; ⁶School of Engineering, University of Glasgow, Glasgow, Scotland, UK.

‡ Equal contribution

*Corresponding author

Mailing address: Nebraska Center for Virology and School of Biological Sciences, University of Nebraska-Lincoln, Morrison Center, 4240 Fair Street, Lincoln, NE 68583-0900

Phone: (402) 472-4550; Fax: (402) 472-3323; E-mail: cwood@unlnotes.unl.edu

† Present address: Ragon Institute of MGH, MIT and Harvard, Boston

Data from this chapter was published in PLoS One 2012; 7(3): e32827
Abstract

Understanding the evolution of the human immunodeficiency virus type 1 (HIV-1) envelope during disease progression can provide tremendous insights for vaccine development, and simian-human immunodeficiency virus (SHIV) infection of nonhuman primate provides an ideal platform for such studies. A newly developed clade C SHIV, SHIV-1157ipd3N4, which was able to infect rhesus macaques, closely resembled primary HIV-1 in transmission and pathogenesis, was used to infect several pig-tailed macaques. One of the infected animals subsequently progressed to AIDS, whereas one remained a nonprogressor. The viral envelope evolution in the infected animals during disease progression was analyzed by a bioinformatics approach using ultra-deep pyrosequencing. Our results showed substantial envelope variations emerging in the progressor animal after the onset of AIDS. These envelope variations impacted the length of the variable loops and charges of different envelope regions. Additionally, multiple mutations were located at the CD4 and CCR5 binding sites, potentially affecting receptor binding affinity, viral fitness and they might be selected at late stages of disease. More importantly, these envelope mutations are not random since they had repeatedly been observed in a rhesus macaque and a human infant infected by either SHIV or HIV-1, respectively, carrying the parental envelope of the infectious molecular clone SHIV-1157ipd3N4. Moreover, similar mutations were also observed from other studies on different clades of envelopes regardless of the host species. These recurring mutations in different envelopes suggest that there may be a common evolutionary pattern and selection pathway for the HIV-1 envelope during disease progression.
Introduction

The envelope gene of human immunodeficiency virus type 1 (HIV-1) is the most genetically diverse among all HIV-1 genes. The vital role of HIV-1 envelope in determining cell tropism of the virus and escape from host immune surveillance made it a logical choice as the main focus for vaccine development. Thus, a better understanding of how the envelope evolves during disease progression could aid in designing better vaccines. Several envelope mutations, such as increases in the length of V1V2 variable loops and number of potential N-glycosylation sites (PNGS), have been linked with disease progression in humans (44, 80, 322). Since these mutations were observed in envelopes from different clades, it would suggest that the envelope might tend to follow a certain evolutionary pattern during disease progression. Infection of non-human primates with simian-human immunodeficiency virus (SHIV) would be an ideal platform for investigating such HIV-1 envelope evolution during disease progression.

SHIV strains have been a significant tool in studying the role of HIV-1 envelope in pathogenesis and the development of AIDS vaccines for over a decade. Since their inception, SHIV constructs have undergone dramatic improvements to recapitulate many of the features of primary HIV-1 infection when used to infect rhesus macaques. One such design, SHIV-1157ipd3N4, expresses an R5 tropic HIV-1 clade C envelope isolated from a Zambian infant (358). In addition, SHIV-1157ipd3N4 is pathogenic and fully capable of mucosal transmission through multiple routes (65, 358). These properties closely resemble those of recently transmitted HIV-1 isolates, which are mostly R5 tropic and transmitted via mucosal routes (69, 92, 279, 298). The fact that SHIV-1157ipd3N4 carries an HIV-1 clade C envelope makes this SHIV an important model to study
transmission and pathogenesis of HIV-1 infection in humans: because more than fifty percent of all HIV-1 infections worldwide are caused by HIV-1 clade C (158, 315).

Until recently, SHIV-1157ipd3N4 had only been utilized to infect rhesus macaques (*Macaca mulatta*), a popular animal model for studying HIV pathogenesis and AIDS vaccine development. However, there is now a heightened interest in the pig-tailed macaque (*Macaca nemestrina*) model. In comparison to rhesus macaques, pig-tailed macaques are unique as they express a defective host restriction factor TRIM5α and hence are susceptible to infection by simian-tropic HIV-1 strains (38, 153, 173, 377). Given this recent focus on the pig-tailed macaque animal model, it will be important to examine the relationship between disease progression and envelope evolution in pig-tailed macaques infected by SHIV-1157ipd3N4. Moreover, this study will help to determine whether envelope changes observed in a human and in a rhesus macaque during disease progression are present in this pig-tailed model. In an earlier report, Ho et al. demonstrated that SHIV-1157ipd3N4 was able to infect pig-tailed macaques by intrarectal inoculation and to cause AIDS in the infected animal (161). We procured samples from these infected pig-tailed macaques and employed 454 ultra-deep pyrosequencing (UDPS) to characterize the HIV-1 clade C envelope changes as the infected animal progressed to AIDS.

UDPS is a high-throughput sequencing technology that can rapidly generate a vast amount of sequencing data in a cost-effective manner. The high sensitivity of this technology in detecting minor populations makes it an ideal approach for our study. UDPS has been used by several groups to study minor drug resistance mutations, cytotoxic T-lymphocytes (CTL) escape pathways and small envelope region such as V3
To our knowledge, this is the first study to use UDPS to resolve HIV-1 clade C envelope evolution in infected pig-tailed macaques during disease progression which is marked by a consistent decline in CD4$^+$ T-cell count and the present of high plasma viral load. AIDS is established when the CD4$^+$ T-cell count falls to <200 cells/ul. By using UDPS, we were able to quantitatively assess the emergence and outgrowth of minor variants in the infected animals as the disease progressed and provide a detailed genetic analysis of such variants. Interestingly, we found several gp120 mutations that evolved in parallel with disease progression. In addition, we compared the mutations observed in the pig-tailed macaque with published data on the evolution of the precursor envelope of SHIV-1157ipd3N4 in a rhesus macaque and the Zambian infant, from whom the viral envelope was originally isolated (387, 438). Our analysis showed that several shared gp120 mutations were prevalent in the infected human, rhesus and pig-tailed macaques throughout disease progression. This study suggests that the HIV-1 clade C envelope may follow a comparable evolutionary pattern and host selective pressure during disease progression in infected hosts, regardless of the host species.
Materials and methods

Animals and viral stocks

All animals used in this study were housed and cared for according to the Guide for the Care and Use of Laboratory Animals at the Washington National Primate Research Center (WaNPRC), an Association for Assessment and Accreditation of Laboratory Animal Care International accredited institution. The animal quarters are maintained at 75–78°F with controlled air humidity and quality. The home cages of the animals are steam cleaned bimonthly and the waste pans are cleaned daily. Commercial monkey chow is fed to the animals once daily and drinking water is available at all times. Daily examination and any medical care of the animals are provided by the veterinary staff of WaNPRC in consultation with the clinical veterinarian. The experimental procedures were approved by the Institutional Animal Care and Use Committee (2370-20) at the University of Washington and conducted in compliance with the Public Health Services Policy on Humane Care and Use of Laboratory Animals (http://grants.nih.gov/grants/olaw/references/PHSPolicyLabAnimals.pdf). The animals were kept under deep sedation during all procedures with ketamine HCl at the dose of 10–15 mg/kg intramuscularly to alleviate any pain and discomfort. The animals were monitored by the Animal Technician or Veterinary Technologist while under sedation.

The construction of the infectious molecular clone, SHIV-1157ipd3N4, and the preparation of the viral stock were described previously (358). All animal procedures and immunological analysis have also been published (161). Briefly, four juvenile pig-tailed macaques were inoculated with SHIV-1157ipd3N4 intrarectally. Infected animals were
monitored over a period of 84 weeks postinoculation. Peripheral blood mononuclear cell (PBMC) and tissue samples were collected from the infected animals periodically.

**PCR amplification and amplicon library preparation for UDPS**

Genomic DNA from PBMC and gut tissue samples was extracted following standard protocols. For amplicon library preparation, the full envelope was amplified from the samples with first round PCR primers positioned outside the envelope gene. The envelope from each sample was further amplified into 6 amplicons with six pairs of primers during the second round PCR. The envelope regions amplified by these primer pairs were V1V2C2 by primers env1, C2V3C3 by primers env2, V3C3V4C4 by primers env3, C4V5C5-gp41 by primers env4, C5-gp41 by primers env5 and gp41 by primers env6. Each sample was barcoded with a specific 10-nucleotides multiplex identifier (MID) and only primers containing the forward adaptor sequences were barcoded (Table 3). Two additional primers, env3_F_control and env3_R_control, were designed to amplify the envelope of an infectious HIV-1 clade C construct (1084ic) which served as a control (149). Primer env1_F was modified to env1_F_a for amplifying the 65 and 84 weeks post-inoculation samples from J02185. Sequences of these primers are presented in Table 3. The PCR conditions used for amplicon amplification were 1 cycle of 95°C for 2 min, 35 cycles of 95°C for 30 sec, 60°C for 30 sec and 72°C for 30 sec and a final extension of 72°C for 4 min. All PCR was carried out with the FastStart High Fidelity PCR system (Roche, Indianapolis, IN). PCR products were isolated with the E.Z.N.A. Gel Extraction Kit (Omega Bio-Tech, Norcross, GA) and purified by the Agencourt AMPure magnetic beads (Beckman Coulter Genomics, Danvers, MA) following the conditions recommended for the Titanium amplicon library preparation (Roche/454 Life
Sciences, Branford, CT). Purified amplicons were quantitated using the Quant-iT PicoGreen assay kit (Invitrogen, Carlsbad, CA) and pooled in equimolar concentration according to the manufacturer’s recommendations. The pooled amplicons were then processed and sequenced on a Genome Sequencer FLX (Roche/454 Life Sciences, Branford, CT) at the Environmental Genomics Core facility (Engencore), Innovista Research District, University of South Carolina, Columbia, SC.

**UDPS bioinformatics analysis**

The initial sequence reactions yielding 574,225 reads that were processed to ensure high quality reads to reduce the typical sequencing errors from 454/Roche pyrosequencing. The data were cleaned by a set of scripts including the following criteria: (i) a perfect match to both the barcode and forward primer, (ii) >100 bases in length, and (iii) no undetermined bases (N). The 454 reads were then separated into samples by amplicons. The flowgrams corresponding to these reads were extracted, truncated at the first noisy signal (174), and then filtered to remove any read where this occurred in the first half of the flowgram. We then applied the AmpliconNoise pipeline to these samples separately using default parameters for GSFLX Titanium data (304). The flowgrams were clustered with the PyroNoise program to remove 454 sequencing errors (303). The forward primer and barcodes were removed from the resulting sequences, prior to their truncation to 400 bp. These were clustered with the SeqNoise program to remove PCR errors (304). Finally, the Perseus *de novo* chimera classifier was applied to screen the sequences for chimeric PCR products which were then removed, giving denoised chimera checked sequences that were used in the following analysis.

Alignments were generated and manually inspected to ensure any remaining variants
with frameshifts or stop codons were removed. After the automated cleanup, a small number of problematic sequences remained that were associated with indels in homopolymer tracts. We modified these errors by deleting the extra base or adding a missing base relative to the inoculum sequence. As a control for 454 sequencing errors, a plasmid containing a subtype C envelope, 1084ic, was amplified and processed in parallel to the samples.

**Sequence diversity and divergence of intra-host virus populations**

Diversity of viral sequences for each time point within each monkey was calculated using the Shannon Diversity Index. As follows

\[ H' = - \sum_{i=1}^{R} p_i \log p_i \]

where \( H' \) is the Shannon Diversity Index, \( R \) is the total number of species encountered and \( p_i \) represents the fraction of the entire population made up of species \( i \) (348). Divergence of viral sequences for each time point within each monkey was calculated as the genetic distance between each sequence and the inoculum strain as calculated in MEGA v5.0 (373).

**Envelope charge analysis**

The charge of the envelope was calculated with AminoTrack™ (241). Briefly, arginine and lysine have a charge of +1, while aspartic acid and glutamic acid have a charge of -1. Statistical analysis was calculated using GraphPad Prism 5 (GraphPad Software, Inc., San Diego, CA).
Results

Infection of pig-tailed macaques with SHIV-1157ipd3N4

Detailed examination of the immunological responses of the infected pig-tailed macaques has been conducted previously (161). Briefly, after intrarectal inoculation of SHIV-1157ipd3N4, systemic infection was achieved in all four juvenile pig-tailed macaques with an average peak plasma viral RNA load \(>7 \times 10^6\) copies/ml by 2 weeks post-inoculation (Fig. 1B). However, two infected animals (M04123 and L03165) died due to unrelated causes during sampling procedures at 2 and 48 weeks post-inoculation, respectively (data not shown). The remaining animals (K03135 and J02185) were monitored over a period of 84 weeks. The plasma viral loads of macaque K03135 increased while its CD4\(^+\) T-cell counts declined as disease progressed (Fig. 1A and B). Macaque K03135 developed AIDS with a CD4\(^+\) T-cell count consistently \(<200\) cells/ul from 20 weeks post-inoculation onwards. Therefore, macaque K03135 was classified as a progressor. Meanwhile, macaque J02185 had normal CD4\(^+\) T-cell counts over the entire study period (Fig. 1A). Plasma viral loads of macaque J02185 fell below the detection limit at 2 weeks after peak viremia and remained low over the next 82 weeks (Fig. 1B). Macaque J02185 was still healthy at the end of the study and was classified as a non-progressor. Both animals maintained a detectable but gradually declining PBMC proviral DNA load throughout the course of the disease (Fig. 1C).

Macaque sample selection and ultra-deep pyrosequencing of viral envelope

In order to gain a better picture of the envelope changes throughout disease progression, only animals that underwent a full 84 week course of observation were
included for UDP. For this purpose, PBMC samples from 5 time points of the progressor macaque K03135 (2, 16, 44, 64 and 83 weeks post-inoculation) and non-progressor macaque J02185 (2, 16, 44, 65 and 84 weeks post-inoculation) were selected for UDP. In addition, it was reported previously that there was an unexpected peak of viral load in the duodenum sample from K03135 at 16 weeks post-inoculation (161). The cause for this sudden elevated viral load in the gut tissue was unclear. To investigate if there was any unique viral population residing within the gut tissue at this time point, additional gut tissue samples from both animals at 16 weeks post-inoculation were also included for UDP. Lastly, the two animals (M04123 and L03165) that died due to unrelated causes during sampling procedures were excluded for analysis as it would be impossible to extrapolate the data from these animals to disease progression.

The current read length on the GS FLX Titanium series is approximately 400 bp. In order to cover the regions of interest in our study, we designed six overlapping PCR amplicons to capture the viral envelope regions (Fig. 2). Using this approach, we were able to sequence almost the entire gp120 and through the transmembrane domain of gp41. From the UDP, we obtained a total of 574,225 reads with a median of 43,601 (range 13,964 to 81,666) reads from each sample. The UDP data were cleaned to remove reads with PCR and UDP artifacts while attempting to retain as many high-quality sequences as possible (see Materials and Methods). During the initial procedure, a median of 26% (range 23%–30%) reads were discarded from each sample. However, after the application of the AmpliconNoise algorithm, this number increased to a median of 43% (range 34%–47%). While this number may seem excessively high, there were
still on average 18,000 reads per sample, ensuring sufficiently high coverage across the envelope to detect minor variants.

**Sequence diversity and divergence of intra-host viral populations**

Selective pressures imposed by the host immune system are the main driving forces shaping the evolutionary dynamics of the viral envelope. To examine the effect of immune selection on envelope evolution, we quantified the levels of diversity and divergence for each amplicon over time in both animals. The non-progressor macaque J02185 had a relatively low level of diversity in the entire envelope throughout the course of observation, with a maximum Shannon diversity index of only 0.72 in amplicon 5 beginning at 16 weeks post-inoculation (Fig. 3A). On the contrary, diversity for the progressor macaque K03135 began to increase at 44 weeks post-inoculation. It reached a peak of 2.05 in amplicon 1 and 2.71 in amplicon 3 by 64 weeks post-inoculation (Fig. 3A). The amplicons that showed a high level of diversity mainly comprised gp120. A relatively low level of diversity occurred within the gp41 in both animals. Similar to the levels of diversity, the nonprogressor macaque J02185 had low divergence, showing that its viral populations did not deviate significantly from the inoculum strain over time, with a highest p-distance of only 0.007 in amplicon 1 at 65 weeks post-inoculation (Fig. 3B). The divergence of progressor macaque K03135 had a pattern similar to its diversity, with a progressively elevated p-distance from 44 weeks post-inoculation onwards. Amplicons 1 and 3 of the progressor macaque K03135 showed a high level of deviation from the inoculum strain, reaching a p-distance of 0.034 and 0.025, respectively, by 84 weeks post-inoculation (Fig. 3B). Additionally, the gp41 of K03135 developed a slightly higher divergence than the other envelope regions at 44 weeks post-inoculation (Fig. 3B).
Mutations in this region might have evolved in response to the host immune pressure or selection of viral fitness, but there is no evidence or literature to support these possibilities. Lastly, the accumulation of minor mutations in gp41 of K03135 did increase its divergence despite a rather homogenous population over time as shown in the diversity. For the plasmid control, each of the six amplicons showed no diversity and a single population was present.

**Envelope evolution through disease progression**

In order to further demonstrate the progressive nature of the diversification of envelope throughout disease progression, we aligned the amplicons from different time points with their corresponding reference sequence of the inoculum envelope. Furthermore, to highlight the similarities of some mutations observed in the infected pig-tailed macaque with other species, we included previous sequence data from a study involving the infected human and rhesus macaque for comparison (387). In agreement with the diversity and divergence analysis, alignment of the amplicons from non-progressor macaque J02185 revealed extremely low variations within each time point and from the inoculum envelope (Fig. S1).

**C1 region**

Contrary to J02185, substantially more mutations were observed in the progressor macaque K03135 as this animal progressed to AIDS. Among the mutations in K03135 was S124P, which evolved from being the minor population at 44 weeks postinoculation to becoming the majority variant by 64 weeks postinoculation (Fig. 4A). S124P is adjacent to the CD4 binding site and a proline mutation at this position might affect the
CD4 binding site structure. Interestingly, the parental envelope of SHIV-1157ipd3N4 also contains a proline at position 124 (172). Besides S124P, we also observed the emergence of K130N from 64 weeks post-inoculation onwards (Fig. 4A). K130N resulted in an additional PNGS and its appearance in the main viral population coincided with a consistently low CD4 count and the development of AIDS. More importantly, this identical mutation was also observed in the human and rhesus macaque during disease progression (Fig. 5A and Table 1) (387). In addition, the same mutation also developed during disease progression in a rhesus macaque infected with a clade B SHIV (165).

**V1V2 loops**

14% and 6.2% of the viral populations contained deletions and insertions in the V1 loop, respectively, at 64 weeks postinoculation (Fig. 4A). By 83 weeks, >80% of the viral populations contained an insertion in V1 but only 18.3% contained deletions. Deletions and insertions in the V1 loop are also common in human and rhesus macaque (Fig. 5A and Table 1) (387). While insertions in the V1 loop usually involved PNGS, many of the mutations in the V1V2 region comprised of charged amino acids. The net charge of the V1V2 region for K03135 decreased significantly over time (P<0.0001) and was strikingly similar to those observed in the rhesus macaque (Fig. 6A). This charge reduction in V1V2 region was only observed as the animals progressed to AIDS.

**C2 region**

Unlike the highly variable V1V2 region, only 3 non-transient mutations evolved within C2 region of Env. Among these mutations, G262E and D273N represent merely 0.9% of the viral populations at 44 weeks post-inoculation, but subsequently emerged as
the dominant viral population by 64 weeks postinoculation (Fig. 4B). These two 
mutations had also been observed in the rhesus macaque during disease progression (Fig. 
5B, Table 1). Interestingly, D273N locates within the CD4 binding site and mutation at 
this position may affect envelope binding to CD4. More importantly, these mutations 
occurred only after the onset of AIDS in both the pig-tailed and rhesus macaques.

**V3 loop**

The V3 loop of Env was relatively conserved in K03135, except for S302H that 
emerged at 44 weeks post-inoculation (Fig. 4B). Histidine at this position is extremely 
rare, 0 out of 756 sequences, in HIV-1 clade C envelope in the HIV sequence database. It 
is more commonly found in envelopes from other clades, such as A, B and D. So far, we 
had only observed 1 out of 18 HIV-1 clade C infected patients from a Zambian cohort 
with histidine at this position (unpublished data). Since small changes in this region of the 
V3 loop could disturb the stability of the envelope trimeric structure of and affect CCR5 
binding, mutations at this position could be important (371, 427). Surprisingly, we 
detected a minor population of 7.4% in the gut tissue of K03135 at 16 weeks 
postinoculation that contained a large deletion comprising the regions of C2V3C3 (Fig. 
4B). It was determined that this was not as result of PCR or UDPS, given that it has 
passed our strict quality control procedure and contained an intact open reading frame. At 
this point, it is not clear whether this viral population is functional. However, it was 
reported previously that an envelope with truncation in V3 can still be functional (8, 277). 
In addition, given that this population was unique to the gut tissue, it might serve as 
another example of HIV-1 compartmentalization (394).
C3 region

The high degree of genetic polymorphism in the α2-helix of C3 had been noted previously in several studies involving human and rhesus macaque (Fig. 5C) (264, 387, 430). The identical region is also highly variable in K03135 from 64 weeks post-inoculation onwards (Fig. 4C). Similar to the V1V2 region, mutations in this region frequently involved charged amino acids. A comparison of the net charge of C3 revealed a close similarity between the pigtailed and rhesus macaques, with a significant (P<0.0001) decrease in the C3 charge occurring near the late stage of disease (Fig. 6B).

V4 loop

85% of the viral populations carried deletions in V4 by 64 weeks post-inoculation (Fig. 4C). Deletions in V4 had been reported in pig-tailed macaque infected with a different SHIV (361). In our study, the V4 deletions focused on the methionine located at the tip of V4, which was also progressively eliminated over time in the precursor envelope sequences of SHIV-1157ipd3N4 (Fig. 5C, Table 2) (387). Surprisingly, deletions in V4 were presented at 16.5% of the viral populations from the non-progressor macaque J02185 at 65 weeks post-inoculation (Fig. S1).

C4 region

Despite a high sequence homology in the C4 region of Env, E415Q and E426A in K03135 emerged from 64 weeks postinoculation onwards (Fig. 4C). E415Q is a significant mutation since this position is involved in CD4 binding. Surprisingly, the identical position was also mutated to glutamine over time in rhesus macaque infected with another SHIV (165). Similarly, E426A, a CCR5 binding site, was observed in pig-
tailed macaques, rhesus macaques and human as the disease progressed (Fig. 5C, Table 2). Since glutamic acid is a negatively charged amino acid, mutations at these two positions increased the overall charge of C4 (P<0.0001) as the disease progressed in K03135 (Fig. 6C). In addition, this pattern of increasing the charge of C4 resembled that observed in rhesus macaque during disease progression (Fig. 6C).

**V5 loop and gp41 region**

91.6% of the viral populations in K03135 contained deletions in V5 at 83 weeks post-inoculation (Fig. 7A). Similar deletions were also observed in rhesus macaque and human over time (Fig. 5D, Table 2). Lastly, the genetic variability of the envelope decreased sharply after the V5 loop. We did not find any mutations in the C5 region of the envelope and only few mutations were detected in gp41 (Fig. 7B and C). However, 1.2% of the viral populations in the gut tissue of K03135 contained deletions in the heptad repeat 1 of gp41 (Fig. 7C). Given that this deletion happened at such low frequency and localized in the gut tissue only, the impact of this mutation had on the overall disease progression could be minimal.
Discussion

In this study, we provided a comprehensive view of the HIV-1 clade C envelope diversifications during disease progression between progressor and non-progressor pig-tailed macaques. Despite an extremely low level of diversity and divergence from the inoculum, we were still able to detect minor envelope variants in the non-progressor macaque J02185 over time, demonstrating the value of UDPS. However, there was clearly a mechanism by which J02185 was able to keep its infection well controlled. The fact that J02185 had envelope binding antibodies but no neutralizing antibodies against SHIV-1157ipd3N4, as reported in the previous study, indicated that neutralizing antibody responses are not the main mechanism behind its control over the virus (161). In addition, the gradual reduction of PBMC proviral load occurred in J02185 without any significant decreases in its CD4+ T-cell counts, suggesting that this reduction was not due to a lack of target cells. This differs from the observation in the progressor macaque K03135, whose reduction of PBMC proviral load coincided with low CD4+ T-cell counts. Together, the data strongly suggests that cell-mediated immunity, such as cytotoxic T lymphocyte (CTL) responses, may be responsible for suppressing viral replication in J02185. However, due to a lack of information on the MHC class and CTL response analysis of J02185, we can only postulate that this was the main mechanism for viral control in this animal, which led to a reduction in virus production and the number of infected cells. Since mutations occur more frequently during viral replication, a diminished virus production will minimize the level of viral genetic variation, thus explaining the lack of envelope diversity and divergence in J02185 during the course of observation.
On the contrary, our data showed a very different picture for the progressor macaque K03135. The high plasma viral RNA load, PBMC proviral DNA load and decreasing CD4\(^+\) T-cell counts before the development of AIDS, indicated a lack of viral control by K03135 host responses. Similar to J02185, there were envelope binding antibodies but no neutralizing antibodies against SHIV-1157ipd3N4 present in K03135 (161). Given this lack of neutralizing antibody responses, there may have been a lack of selective pressure resulting in rapid envelope evolution during the early phase of infection, thus explaining the low level of envelope diversity and divergence in K03135 before the development of AIDS. At the late stage of the disease, perhaps due to the combined effects of an ablated cellular immune system and natural selection for viral fitness, there was then a sharp increase in the envelope diversity and divergence in K03135. The notion that viral fitness was the main component for natural selection at the late stage of disease is supported by our observation that the distribution of envelope variations is not random; with mutations frequently occurring at specific regions or hotspots throughout the envelope that are proximal to important receptor binding sites over time, suggesting the presence of an active selection process.

In our study, we also observed that the envelope tended to follow a certain evolutionary pattern that correlated with disease progression. The majority of the envelope mutations were observed after the onset of AIDS. The most common mutations involved the addition of PNGS, which usually takes place in highly variable regions such as V1V2, \(\alpha2\)-helix of C3, V4 and V5, and PNGS have been shown to play an important role in immune evasion (70, 405). However, since the infected animals in our study had no neutralizing antibodies against the homologous virus, we believe these mutations
evolved in response to the presence of high levels of non-neutralizing antibodies (161). Although non-neutralizing antibodies cannot inhibit viral infection directly, they might still exert a selective pressure on the viral envelope through antibody-dependent cell-mediated cytotoxicity (ADCC), as suggested recently during SIV infection (370).

Changes in the length of envelope had been associated with immune escape and disease progression (80, 392). Our data show that V1 accounted for the largest increase in length over time, and coincided with late stages of disease in the infected animals. There is a close relationship between the higher percentages of viral populations containing deletions in V4 and V5 with late stages of disease as well. Furthermore, our study observed a close association of changes in the envelope charges with disease progression. For example, the net charge in V1V2 and C3 tended to increase, while there was a decrease in C4 charges as the disease progressed. Variations in the charge of envelope had been suggested to affect viral fitness and were associated with disease progression (311). The high number of charged amino acids in the α2-helix of C3 and its close proximity to the V4 loop suggest that it could be under heavy selective pressure from the non-neutralizing antibodies (146, 264, 265). More importantly, charged amino acid mutations such as D273N, E415Q and E426A might affect the receptors binding ability of the envelope. Interestingly, similar mutations had been observed with clade A envelope as well (96).

In our study, we have demonstrated that the use of SHIV-1157ipd3N4 in pig-tailed macaque model can mimic the primary HIV-1 infection and disease progression in the human. Due to samples availability, we only analyzed PBMC vDNA. While plasma vRNA is derived from actively replicating virus population, infected PBMC contain
incoming or newly synthesized vRNA in addition to integrated and unintegrated vDNA. Thus, studying PBMC allows us to document the complete HIV quasispecies repertoire which consists of actively replicating and archived viruses. Moreover, the use of UDPS in this study allowed us to detect and quantify minority variants that would have gone unnoticed using conventional sequencing technology. However, a major challenge with UDPS is its inherent errors that arise during the pyrosequencing process and those introduced by PCR amplification. Such errors are typically localized to homopolymers (397). Therefore, careful filtering and control is essential to minimize the possibility of erroneous sequences. We implemented a carefully designed clean-up strategy to minimize the impact of 454 sequencing errors on interpreting our data. Our denoising approach removed most pyrosequencing errors due to sporadic base changes introduced during PCR while attempting to retain true biological sequence variation. The success of this approach is measured by our control plasmid amplicons as they were reduced to a single variant population.

In summary, UDPS has shown a strong association between the mutational dynamics of envelope and disease progression in the infected animals. In addition, we found a temporal relationship between the numbers of mutations occurring with little genetic variation before AIDS and more genetic variation afterwards. Such changes are likely due to a lack of immune surveillance in the early phase and selection for better viral fitness at the late stage of disease. Additionally, our data suggest that although nonneutralizing antibodies cannot inhibit the virus directly, they might still contribute to the evolution of envelope. More importantly, a majority of the mutations and evolutionary patterns of the envelope witnessed in SHIV-1157ipd3N4-infected pig-tailed
macaques were also common in infected rhesus macaques and human, thus, suggesting a common selection pathway for the virus irrespective of the species studied.
Acknowledgments

We thank Ben Bimber and Bette Korber for useful discussions on the 454 analysis.

Author Contributions

Conceived and designed the experiments: FYT DCT CW. Performed the experiments: SG OH PP. Analyzed the data: FYT DCT CQ. Contributed reagents/materials/analysis tools: SLH RMR. Wrote the paper: FYT DCT.
Figures and Figure Legends

Figure 1.

A

\[ 1 \times 10^4 \]

\[ 1 \times 10^3 \]

\[ 1 \times 10^2 \]

\[ 1 \times 10^1 \]

-1 0 1 2 3 4 5 6 7 8 9

Weeks post-inoculation

CD4 T cells (cells/μl)

J02185

K03135

B

\[ 1 \times 10^8 \]

\[ 1 \times 10^7 \]

\[ 1 \times 10^6 \]

\[ 1 \times 10^5 \]

\[ 1 \times 10^4 \]

\[ 1 \times 10^3 \]

\[ 1 \times 10^2 \]

-1 0 1 2 3 4 5 6 7 8 9

Weeks post-inoculation

vRNA copies/ml of plasma

J02185

K03135

C

\[ 1 \times 10^4 \]

\[ 1 \times 10^3 \]

\[ 1 \times 10^2 \]

\[ 1 \times 10^1 \]

\[ 1 \times 10^0 \]

-1 0 1 2 3 4 5 6 7 8 9

Weeks post-inoculation

Proval DNA copies/μg of DNA

J02185

K03135
Figure 1. CD4$^+$ T-cell counts, plasma and PBMC viral loads from infected pig-tailed macaques. (A) Total CD4$^+$ T-cell counts (B) viral RNA loads in plasma and (C) proviral cDNA loads in PBMC. A scale break (/) representing 8 weeks post-inoculation was insert into the x-axis to show the early stage of infection. A cross (+) represents death of the animal due to AIDS.
Figure 2.
Figure 2. Schematic representation of the envelope regions covered by each amplicon (in reference to Env of the infectious molecular clone, SHIV-1157ipd3N4).
Figure 3.
Figure 3. Phylogenetic analysis. (A) Shannon diversity. Longitudinal trends in sequence diversity analyzed using the Shannon Index. Time post-inoculation is shown on the x-axis and Shannon Diversity Index values are shown on the y-axis. All data shown are for AmpliconNoise processed amplicons. (B) Divergence. Longitudinal trends in sequence divergence from the time of inoculation. For direct comparison of sequences the mean Hamming distance at each time point was measured. Divergence was based on the Hamming distances to the SHIV-1157ipd3N4 inoculum for each amplicon. Time post-inoculation is shown on the x-axis and Hamming distance values are shown on the y-axis. Data for each amplicon are shown separately with macaque J02185 in red and macaque K03135 in blue, respectively.
### Amplicon 1

**Figure 4.**

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (100%)</th>
<th>K44ws_0 (100%)</th>
<th>K16ws_0 (100%)</th>
<th>K16ws_gus_0 (100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>124</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>126</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>127</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Amplicon 2

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>231</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>232</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>233</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>234</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>235</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### B

**Figure 4.**

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>321</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>322</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>323</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>324</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>325</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>431</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>432</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>433</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>434</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>435</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C2

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>541</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>542</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>543</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>544</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>545</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C3

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>651</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>652</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>653</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>654</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>655</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C4

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>761</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>762</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>763</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>764</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>765</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C5

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>871</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>872</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>873</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>874</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>875</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C6

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>981</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>982</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>983</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>984</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>985</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C7

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1091</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1092</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1093</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1094</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1095</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C8

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1201</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1202</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1203</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1204</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1205</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C9

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2301</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2302</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2303</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2304</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2305</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C10

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3401</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3402</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3403</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3404</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3405</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C11

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4501</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4502</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4503</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4504</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4505</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C12

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5601</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5602</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5603</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5604</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5605</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### C13

<table>
<thead>
<tr>
<th>SHIV-1157ipDIN4</th>
<th>K42ws_0 (110%)</th>
<th>K44ws_0 (110%)</th>
<th>K16ws_0 (110%)</th>
<th>K16ws_gus_0 (110%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6701</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6702</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6703</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6704</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6705</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4. Amino acid alignments of the envelope from infected pig-tailed macaque K03135 as represented by ‘‘K’’. (A) Amplicon 1, (B) Amplicon 2 and (C) Amplicon 3. Weeks post-inoculation is represented by ‘‘wks’’. The value after ‘‘_’’ shows the population number. Deletions in the alignments are shown as ‘‘-’’. The amount of the particular viral population at that time point is represented as percentage. Point mutations described in the text are highlighted by red color box.
Figure 5. Amino acid alignments of the envelope from infected human (1157i) and rhesus macaque (RPn-8). 1157i and RPn-8 examples for (A) Amplicon 1, (B) Amplicon 2, (C) Amplicon 3 and (D) V5 loop. The inoculum for RPn-8 was SHIV-1157i (the initial infectious molecular clone). Weeks post-inoculation or infection is represented by ‘‘wks’’. Sequences presented here are examples from each time point. These are not consensus sequences and do not represent all the sequence data from both the infected human and rhesus macaque. Point mutations described in the text are highlighted by red color box.
Figure 6.

A

K03135

RPn-8

P < 0.0001

Charges of V1V2

SHIV-1157/pRDN4

2 wpi 16 wpi 44 wpi 64 wpi 83 wpi

26 wpi 86 wpi 159 wpi 215 wpi 275 wpi

B

K03135

RPn-8

P < 0.0001

Charges of C3 region

SHIV-1157/pRDN4

2 wpi 16 wpi 44 wpi 64 wpi 83 wpi

26 wpi 86 wpi 159 wpi 215 wpi 275 wpi

C

K03135

RPn-8

P < 0.0001

Charges of C4 region

SHIV-1157/pRDN4

2 wpi 16 wpi 44 wpi 64 wpi 83 wpi

26 wpi 86 wpi 159 wpi 215 wpi 275 wpi

Time points post-inoculation
Figure 6. Comparison of envelope charges between the infected pig-tailed macaque (K03135) and rhesus macaque (RPn-8). (A) V1V2 regions, (B) C3 region and (C) C4 region. Time points are represented in weeks post-inoculation (wpi). The inoculum SHIV strain for K03135 was SHIV-1157ipd3N4 and SHIV-1157i for RPn-8. Each icon represents one envelope sequence.
### Figure 7.

#### Amplicon 4

<table>
<thead>
<tr>
<th></th>
<th>C4</th>
<th>V5</th>
<th>C5</th>
<th>gp41</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHIV-1157lpd3N4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K2wks 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K16wks 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K15wks_out 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K44wks 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R44wks 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R44wks 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K6wks 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K3wks 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K3wks 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K8wks 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Amplicon 5

<table>
<thead>
<tr>
<th></th>
<th>C5</th>
<th>gp41</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHIV-1157lpd3N4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K2wks 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K16wks 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K15wks_out 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K16wks_out 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K44wks 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K44wks 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R44wks 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K6wks 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K6wks 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K6wks 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K8wks 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K8wks 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## C

### (Amplicon 6)

| SHIV-1157ipd3N4 | QARVLAIERY | LQGQQLG6W | GC54XKLCYTT | AVPWNSWSWN | KSQUDIWDSW | TSNQYDREIS | RSTDDTVEEL | EDGQMQQORM | EKQDLALBSW | KNLWNNPSYT | RKLWYIKIPI | MIVGSLIGLR | IIFAVLSIVW | RV |
|-----------------|-------------|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-----------|-------------|-----|
| K2wks_0         | (100%)      |            |             |             |             |             |             |             |             |             |             |             |           |             |     |
| K16wks_0        | (100%)      |            |             |             |             |             |             |             |             |             |             |             |           |             |     |
| K16wks_gut_0    | (100%)      |            |             |             |             |             |             |             |             |             |             |             |           |             |     |
| K44wks_2        | (0.1%)      | M.         |             |             |             |             |             |             |             |             |             |             |           |             |     |
| K44wks_0        | (38.1%)     |            |             |             |             |             |             |             |             |             |             |             |           |             |     |
| K44wks_1        | (61.7%)     | V.         | M.          |             |             |             |             |             |             |             |             |             |           |             |     |
| K64wks_0        | (100%)      | V.         | M.          |             |             |             |             |             |             |             |             |             |           |             |     |
| K83wks_0        | (99.8%)     | V.         | M.          |             |             |             |             |             |             |             |             |             |           |             |     |
| K83wks_1        | (0.2%)      | V.         | M.          |             |             |             |             |             |             |             |             |             |           |             |     |

| gp41 |
Figure 7. Amino acid alignments of the envelope from infected pig-tailed macaque K03135 as represented by ‘‘K’’. (A) Amplicon 4, (B) Amplicon 5 and (C) Amplicon 6. Weeks post-inoculation is represented by ‘‘wks’’. The value after ‘‘_’’ shows the population number. Deletions in the alignments are shown as ‘‘-’’. The amount of the particular viral population at that time point is represented as percentage.
<table>
<thead>
<tr>
<th>Time points (post inoculation or infection)</th>
<th>Amino acid position 124 (C1)</th>
<th>Amino acid position 130 (C1)</th>
<th>Deletion (V1)</th>
<th>Insertion (V1)</th>
<th>Amino acid position 262 (C2)</th>
<th>Amino acid position 273 (C2)</th>
<th>Amino acid position 302 (V3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHIV-1157ipd3N4</td>
<td>S (100%)</td>
<td>K (100%)</td>
<td>0%</td>
<td>0%</td>
<td>G (100%)</td>
<td>D (100%)</td>
<td>S (100%)</td>
</tr>
<tr>
<td>K03135 2 weeks</td>
<td>S (100%)</td>
<td>K (100%)</td>
<td>0%</td>
<td>0%</td>
<td>G (100%)</td>
<td>D (100%)</td>
<td>S (100%)</td>
</tr>
<tr>
<td>K03135 16 weeks</td>
<td>S (100%)</td>
<td>K (100%)</td>
<td>0%</td>
<td>0%</td>
<td>G (100%)</td>
<td>D (100%)</td>
<td>S (100%)</td>
</tr>
<tr>
<td>K03135 16 weeks gut</td>
<td>S (100%)</td>
<td>K (100%)</td>
<td>0%</td>
<td>0%</td>
<td>G (100%)</td>
<td>D (92.6%)</td>
<td>S (92.6%)</td>
</tr>
<tr>
<td>K03135 44 weeks</td>
<td>S (84.2%)/P (15.8%)</td>
<td>K (100%)</td>
<td>0%</td>
<td>0%</td>
<td>G (99.1%)/E (0.9%)</td>
<td>D (99.1%)/N (0.9%)</td>
<td>H (41.6%)</td>
</tr>
<tr>
<td>K03135 64 weeks</td>
<td>P (100%)</td>
<td>K (19.5%)/N (80.5%)</td>
<td>14%</td>
<td>6.2%</td>
<td>G (2.9%)/E (97.1%)</td>
<td>D (2.9%)/N (97.1%)</td>
<td>H (100%)</td>
</tr>
<tr>
<td>K03135 83 weeks</td>
<td>P (100%)</td>
<td>N (100%)</td>
<td>18.3%</td>
<td>81.3%</td>
<td>E (100%)</td>
<td>N (100%)</td>
<td>H (100%)</td>
</tr>
<tr>
<td>1157i 0 week</td>
<td>–</td>
<td>K (100%)</td>
<td>0%</td>
<td>0%</td>
<td>R (96.8%)/W (3.2%)</td>
<td>D (100%)</td>
<td>R (100%)</td>
</tr>
<tr>
<td>1157i 26 weeks</td>
<td>–</td>
<td>K (82.7%)/N (6.8%)</td>
<td>37.9%</td>
<td>55.1%</td>
<td>R (100%)</td>
<td>D (100%)</td>
<td>R (100%)</td>
</tr>
<tr>
<td>1157i 125 weeks</td>
<td>–</td>
<td>K (19.2%)/N (46.1%)</td>
<td>0%</td>
<td>100%</td>
<td>R (26.9%)/G (73.1%)</td>
<td>D (100%)</td>
<td>R (100%)</td>
</tr>
<tr>
<td>1157i 155 weeks</td>
<td>–</td>
<td>K (13.6%)/N (81.8%)</td>
<td>9%</td>
<td>81.8%</td>
<td>R (4.5%)/G (95.5%)</td>
<td>D (100%)</td>
<td>R (100%)</td>
</tr>
<tr>
<td>1157i 206 weeks</td>
<td>–</td>
<td>N (100%)</td>
<td>0%</td>
<td>100%</td>
<td>G (100%)</td>
<td>D (100%)</td>
<td>R (100%)</td>
</tr>
<tr>
<td>1157i 288 weeks</td>
<td>–</td>
<td>K (15.6%)/N (18.7%)/T (65.7%)</td>
<td>0%</td>
<td>100%</td>
<td>R (81.2%)/G (18.8%)</td>
<td>D (100%)</td>
<td>R (65.6%)/K (31.2%)/G (3.1%)</td>
</tr>
<tr>
<td>SHIV-1157i</td>
<td>P (100%)</td>
<td>K (100%)</td>
<td>0%</td>
<td>0%</td>
<td>R (100%)</td>
<td>D (100%)</td>
<td>R (100%)</td>
</tr>
<tr>
<td>RpN8 26 weeks</td>
<td>–</td>
<td>K (100%)</td>
<td>0%</td>
<td>0%</td>
<td>R (100%)</td>
<td>D (100%)</td>
<td>R (100%)</td>
</tr>
<tr>
<td>RpN8 86 weeks</td>
<td>–</td>
<td>K (100%)</td>
<td>64.5%</td>
<td>0%</td>
<td>R (100%)</td>
<td>D (96.8%)/G (3.2%)</td>
<td>R (80.6%)/A (16.1%)/S (3.2%)</td>
</tr>
<tr>
<td>RpN8 159 weeks</td>
<td>–</td>
<td>K (100%)</td>
<td>3.3%</td>
<td>0%</td>
<td>R (23.3%)/G (26.7%)/E (50%)</td>
<td>D (100%)</td>
<td>S (100%)</td>
</tr>
<tr>
<td>RpN8 215 weeks</td>
<td>–</td>
<td>K (10%)/N (86.6%)</td>
<td>60%</td>
<td>33.3%</td>
<td>K (3.3%)/G (3.3%)/E (93.3%)</td>
<td>D (40%)/N (60%)</td>
<td>S (100%)</td>
</tr>
<tr>
<td>RpN8 275 weeks</td>
<td>–</td>
<td>N (88.8%)</td>
<td>0%</td>
<td>100%</td>
<td>E (100%)</td>
<td>D (2.8%)/N (97.2%)</td>
<td>S (100%)</td>
</tr>
</tbody>
</table>

**Notes:**
- Adjacent to CD4 binding site
- PNGS and charges

Amino acid positions are in reference to the SHIV-1157ipd3N4 envelope. Percentage of viral population containing a particular mutation is shown. "--" represent no information.
## Table 2. Comparison of HIV-1 clade C Env mutations between infected pig-tailed (K03135), rhesus macaque (RPn-8) and human (1157i).

<table>
<thead>
<tr>
<th>Time points (post inoculation or infection)</th>
<th>Deletion (C2V3C3)</th>
<th>α2-helix (C3)</th>
<th>Deletion (V4)</th>
<th>Amino acid position 415 (C4)</th>
<th>Amino acid position 426 (C4)</th>
<th>Deletion (V5)</th>
<th>Deletion (gp41)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHIV-1157ipd3N4</td>
<td>0%</td>
<td>Non variable</td>
<td>0%</td>
<td>F (100%)</td>
<td>F (100%)</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>K03135 2 weeks</td>
<td>0%</td>
<td>Non variable</td>
<td>0%</td>
<td>E (100%)</td>
<td>E (100%)</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>K03135 16 weeks</td>
<td>0%</td>
<td>Non variable</td>
<td>0%</td>
<td>E (100%)</td>
<td>E (100%)</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>K03135 16 weeks gut</td>
<td>7.4%</td>
<td>Non variable</td>
<td>0%</td>
<td>E (100%)</td>
<td>E (100%)</td>
<td>0%</td>
<td>1.2%</td>
</tr>
<tr>
<td>K03135 44 weeks</td>
<td>0%</td>
<td>Non variable</td>
<td>0%</td>
<td>E (78.8%)/Q (21.2%)</td>
<td>E (86.3%)/A (13.7%)</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>K03135 64 weeks</td>
<td>0%</td>
<td>Variable</td>
<td>85%</td>
<td>E (2.3%)/Q (97.7%)</td>
<td>A (100%)</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>K03135 83 weeks</td>
<td>0%</td>
<td>Variable</td>
<td>100%</td>
<td>Q (100%)</td>
<td>A (100%)</td>
<td>91.6%</td>
<td>0%</td>
</tr>
<tr>
<td>1157i 0 week</td>
<td>0%</td>
<td>Non variable</td>
<td>0%</td>
<td>E (100%)</td>
<td>E (100%)</td>
<td>0%</td>
<td>–</td>
</tr>
<tr>
<td>1157i 26 weeks</td>
<td>0%</td>
<td>Non variable</td>
<td>37.9%</td>
<td>E (96.6%)/G (3.4%)</td>
<td>E (100%)</td>
<td>0%</td>
<td>–</td>
</tr>
<tr>
<td>1157i 125 weeks</td>
<td>0%</td>
<td>Variable</td>
<td>46.2%</td>
<td>E (100%)</td>
<td>E (26.9%)/A (73.1%)</td>
<td>0%</td>
<td>–</td>
</tr>
<tr>
<td>1157i 155 weeks</td>
<td>0%</td>
<td>Variable</td>
<td>95.5%</td>
<td>E (100%)</td>
<td>E (59.1%)/A (40.9%)</td>
<td>0%</td>
<td>–</td>
</tr>
<tr>
<td>1157i 206 weeks</td>
<td>0%</td>
<td>Variable</td>
<td>84.6%</td>
<td>E (100%)</td>
<td>E (15.4%)/A (84.6%)</td>
<td>23.1%</td>
<td>–</td>
</tr>
<tr>
<td>1157i 288 weeks</td>
<td>0%</td>
<td>Variable</td>
<td>12.5%</td>
<td>E (100%)</td>
<td>E (59.4%)/A (40.6%)</td>
<td>15.6%</td>
<td>–</td>
</tr>
<tr>
<td>SHIV-1157i</td>
<td>0%</td>
<td>Non variable</td>
<td>0%</td>
<td>E (100%)</td>
<td>E (100%)</td>
<td>0%</td>
<td>–</td>
</tr>
<tr>
<td>RPn8 26 weeks</td>
<td>0%</td>
<td>Non variable</td>
<td>0%</td>
<td>E (100%)</td>
<td>E (100%)</td>
<td>0%</td>
<td>–</td>
</tr>
<tr>
<td>RPn8 86 weeks</td>
<td>0%</td>
<td>Non variable</td>
<td>0%</td>
<td>E (100%)</td>
<td>E (29%)/A (71%)</td>
<td>0%</td>
<td>–</td>
</tr>
<tr>
<td>RPn8 159 weeks</td>
<td>0%</td>
<td>Variable</td>
<td>60.0%</td>
<td>E (100%)</td>
<td>A (100%)</td>
<td>6.7%</td>
<td>–</td>
</tr>
<tr>
<td>RPn8 215 weeks</td>
<td>0%</td>
<td>Variable</td>
<td>10.0%</td>
<td>E (100%)</td>
<td>A (100%)</td>
<td>83.3%</td>
<td>–</td>
</tr>
<tr>
<td>RPn8 275 weeks</td>
<td>0%</td>
<td>Variable</td>
<td>88.9%</td>
<td>E (100%)</td>
<td>A (100%)</td>
<td>25.0%</td>
<td>–</td>
</tr>
</tbody>
</table>

**Notes:**
- PENGs and charges
- CD4 binding site
- CCR5 binding site
- PENGs and charges
- Heptad Repeat 1

Amino acid positions are in reference to the SHIV-1157ipd3N4 envelope.
Percentage of viral population containing a particular mutation is shown. “-” represent no information.
Table 3. Primers used for ultradeep pyrosequencing library preparation.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)¹</th>
<th>Genome location</th>
</tr>
</thead>
<tbody>
<tr>
<td>env1_F</td>
<td>cgtagcctctccctgcgcactcag-MID-GATGCAATGAGTATAATCAGTATGGGA</td>
<td>309–338²</td>
</tr>
<tr>
<td>env1_F_a</td>
<td>cgtagcctctctctgcgcactcag-MID-CCGGAAAGATAGTATGCTTCACAAAGGG</td>
<td>404–432²</td>
</tr>
<tr>
<td>env1_R</td>
<td>ctatgcgctttgcagccggtcactag-ACATTTGCGATGCTCTGATCTACA</td>
<td>684–707²</td>
</tr>
<tr>
<td>env2_F</td>
<td>cgtagcctctctgtgcgcactcag-MID-CTCTGCTGTTATGCGATATCATAA</td>
<td>640–662²</td>
</tr>
<tr>
<td>env2_R</td>
<td>ctatgcgctttgcagccggtcactag-CAATAGAGAAATTTCTCTCTCAATTTAA</td>
<td>1102–1130²</td>
</tr>
<tr>
<td>env3_F</td>
<td>cgtagcctctctgcgcactcag-MID-ACCAGACAGCAATCTGACCTAGCCAC</td>
<td>912–935²</td>
</tr>
<tr>
<td>env3_R</td>
<td>ctatgcgctttgcagccggtcactag-TCCTCCAGGCCTGAATGTTCTGT</td>
<td>1357–1380²</td>
</tr>
<tr>
<td>env3_F_control</td>
<td>cgtagcctctctctgcgcactcag-MID-ATGAGGATAGGACAGGACAAGCA</td>
<td>892–915²</td>
</tr>
<tr>
<td>env3_R_control</td>
<td>ctatgcgctttgcagccggtcactag-TCCTCCAGGTCTGAAATATCTCTGT</td>
<td>1321–1344²</td>
</tr>
<tr>
<td>env4_F</td>
<td>cgtagcctctctctgcgcactcag-MID-TAACATGGCAGGAGGTTAGGAC</td>
<td>1229–1252²</td>
</tr>
<tr>
<td>env4_R</td>
<td>ctatgcgctttgcagccggtcactag-ATTGAGTCCTGCCTCCCTAGTCT</td>
<td>1546–1568²</td>
</tr>
<tr>
<td>env5_F</td>
<td>cgtagcctctctctgcgcactcag-MID-TGGAGGAGGAGATGAGGGACAA</td>
<td>1374–1397²</td>
</tr>
<tr>
<td>env5_R</td>
<td>ctatgcgctttgcagccggtcactag-TCCTCAACTGCTGCCATTTTCAAGGA</td>
<td>1787–1810²</td>
</tr>
<tr>
<td>env6_F</td>
<td>cgtagcctctctctgcgcactcag-MID-TGGAGGATAGGACAGGAGGCTC</td>
<td>1675–1692²</td>
</tr>
<tr>
<td>env6_R</td>
<td>ctatgcgctttgcagccggtcactag-CGATAATGGATGATATGCGGCTC</td>
<td>2089–2112²</td>
</tr>
</tbody>
</table>

¹Lower-case letters: 5’ extensions for forward and reverse adaptor sequences; MID: 10-bp multiplex identifier; capital letters: envelope specific sequences.
²In reference to the envelope of the infectious molecular clone SHIV-1157ipd3N4 (GenBank: DQ779174.1).
³In reference to the envelope of the HIV-1 isolate HIV1084i (GenBank: AY805330.1).
Supporting Figure 1.
Supporting Figure 1. Amino acid alignments of the envelope from infected pig-tailed macaque J02185 as represented by “J”. (A) Amplicon 1, (B) Amplicon 2, (C) Amplicon 3, (D) Amplicon 4, (E) Amplicon 5 and (F) Amplicon 6. Weeks post-inoculation is represented by “wks”. The value after “_” shows the population number. Deletions in the alignments are shown as “-”. The amount of the particular viral population at that time point is represented as percentage.
Chapter 5

Concluding remarks

In summary, our studies have validated the experimental infection of macaques with SHIV-C as a useful and valid animal model for studying HIV-1 envelope evolution during disease progression. This is supported by our data in chapter two, which has demonstrated that the genetic evolution of HIV-1 envelope V1-V5 region and its disease progression in the SHIV-C infected animal are comparable to those of the HIV-1 infected patient. To further support this conclusion, evidence from chapter four showed similar envelope mutations and evolution in the pig-tailed macaque infected by a closely related SHIV-C. Together, these studies suggest that the HIV-1 envelope V1-V5 region follows a similar evolutionary pattern during disease progression in three different host species.

Although we had observed numerous envelope mutations in chapters two and four based on sequence analysis, the question of whether these mutations had actually affected the envelope functions remained unclear. To address this issue, we examined changes in biological functions of the HIV-1 envelope during disease progression in chapter three. By using the SHIV-C model, we have demonstrated the close relationship between changes in the envelope functions and disease progression. Importantly, we believe are the first to report the occurrence of a reduction in the virion-associated gp120 content as the disease progresses. Moreover, our data also underscored the emergence of mAb neutralization resistance phenotypes with variations in charged residues of the HIV-1 envelope over the disease course. These observations might represent another facet of the viral escape mechanisms aside from the previously proposed glycan shield theory (405).
In addition, data from chapter three suggests that the current notion that the late virus might have a higher fitness in comparison to the early viruses may not necessarily be occurring in all cases during disease progression (310, 385). In contrast with other studies, our results suggest that viruses encoding envelopes from the late disease course could also have relatively lower replicative fitness compare to those viruses with early envelopes. This could be a result of accumulated mutations during disease progression, which may have provided some *in-vivo* survival advantages but at the cost of viral fitness in the late envelopes.

Further studies are warranted to confirm these observations using SHIV encoding other HIV-1 envelope strains. It will be ideal to study these envelope functional changes in HIV-1 patients during disease progression. However, it is infeasible to follow HIV-1 patients over extended period of time from the establishment of infection to the development of AIDS. The current antiretroviral drugs are effective in preventing the disease progression and improving the quality of life in most HIV-1 infected individuals. However, the availability of HAART also presented a major challenge for enrolling drug-treatment naïve patients. These issues represent some of the difficulties in conducting such studies in patients and underscored the importance of non-human primate models for HIV-1 research.

As mentioned in the literature review, there are other envelope functions that we have not investigated. These functions, such as the cytotoxicity and by-stander killing effects, will be interesting to investigate since they might affect the viral pathogenesis. Also, future studies should consider examining the full HIV-1 gp160 instead of only the V1-V5 region, since other envelope regions, such as the signal peptide and gp41, as well
as other viral structural and regulatory gene products might influence the viral fitness and disease progression (17).

Additionally, it is important to note that we cannot exclude the possibility that our observed changes in the HIV-1 envelope over time as an adaptational consequence of infecting non-human primates. A recent study had indeed reported mutations of the HIV-1 clade A envelope that are associated with adaptation during infection of monkey cell lines (173). Lastly, the age of the host during establishment of infection might impact viral evolution as well. The immature immune system of the infants might had exerted less selective pressure during the establishment of infection, in comparison to the adults with mature immune system, resulting in different outcome of selection.

In conclusion, our findings have provided strong evidence that the HIV-1 envelope V1-V5 region follows a similar evolutionary pattern in two different non-human primate species and is closely associated with disease progression in the infected patient, contributing to a better overall understanding of HIV-1 envelope evolution during disease progression.
References


47. **Cale, E. M., P. Hraber, E. E. Giorgi, W. Fischer, T. Bhattacharya, T. Leitner, W. W. Yeh, C. Gleasner, L. D. Green, C. S. Han, B. Korber, and N. L. Letvin.** 2011. Epitope-specific CD8+ T lymphocytes cross-recognize mutant simian immunodeficiency virus (SIV) sequences but fail to contain very early


HIV transmission in resource-poor countries: translating research into policy and practice. JAMA 283:1175-82.


changes in the V3 loop of human immunodeficiency virus type 2 determines the coreceptor usage for CXCR4 and CCR5. Virology 264:237-43.


immunodeficiency virus type 1 envelope and disease progression in early-stage infection. J Infect Dis 178:1306-16.


disease display enhanced viral fitness in parallel with increased gp120 net charge. Virology 379:125-34.


341. **Schupbach, J., M. Popovic, R. V. Gilden, M. A. Gonda, M. G. Sarngadharan, and R. C. Gallo.** 1984. Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. Science 224:503-5.


371. **Suphaphiphat, P., M. Essex, and T. H. Lee.** 2007. Mutations in the V3 stem versus the V3 crown and C4 region have different effects on the binding and fusion steps of human immunodeficiency virus type 1 gp120 interaction with the CCR5 coreceptor. Virology 360:182-90.


181


CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. Nature 384:179-83.


