Adaptive Evolution of the SIV Envelope Protein During Early SIV Infection

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493399

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Adaptive Evolution of the SIV Envelope Protein during Early SIV Infection

A dissertation presented

by

Sergio Ita

ton

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Virology

Harvard University

Cambridge, Massachusetts

January 2016
Adaptive Evolution of the SIV Envelope Protein during Early SIV infection

Abstract

Primate lentiviruses (PLVs), including human immunodeficiency virus type 1 (HIV-1), HIV type 2 (HIV-2), and the simian immunodeficiency viruses (SIVs), cause persistent lifelong infections despite the presence of virus-specific adaptive immune responses. The target of antibodies is the viral envelope glycoprotein (Env), which is expressed on the surface of virions and infected cells as a trimer of gp120:gp41 heterodimers. SIV env sequence variation arising from evasion of antibodies is well established and closely mimics the pattern observed in HIV-infected human patients, yet despite the experimental advantages of the macaque model, the viral dynamics during acute and early infection leading to escape within env have not been well defined. To examine the evolutionary dynamics of SIV env during early infection (≤6 months post-infection), we obtained longitudinal plasma samples from a small cohort of SIVmac251-infected rhesus macaques. We deep-sequenced the env gene from longitudinal samples spanning acute and early infection (2-29 weeks post-infection) from four SIV-infected rhesus macaques and the inoculum (a swarm stock of uncloned SIVmac251) that was used to infect the cohort (by repeated low-dose challenge). Using high-resolution next generation sequencing (NGS), we captured a population bottleneck at the point of transmission from the stock into each animal, tracked the subsequent emergence of Env diversity from the initially homogeneous population, and correlated changes with the onset of Env-specific antibodies. We identified a pattern of common substitutions, insertions, and deletions in env of animals with antibody responses, which repeatedly emerge in SIV-infected primate hosts, and found that adaptive changes clustered within short regions of the V1 and V4 loops of gp120. Surprisingly, we found multiple large in-frame deletions in V4 emerge to become dominant in the viral population in two animals with detectable antibody responses to Env. Furthermore, we developed a novel deep sequencing based viral fitness assay (Fit-Seq) and measured the relative fitness of several key in vivo
antibody escape adaptations. Using Fit-Seq, we found that Env adaptations in V1 and V4 that confer antibody escape had no associated fitness costs in the absence of antibody, but rather, replicated to the same level as SIVmac239 WT. Surprisingly, we also found that Fit-Seq was able to detect antibody-mediated neutralization of SIVmac239 even in cases where activity was undetectable by standard neutralization assays. Our observation of reproducible patterns of Env variation clustered in V1 and V4, together with measurements of relative fitness, suggest that antibody responses can select for mutations that confer viral escape yet have little or no associated replicative fitness cost in the absence of antibody, even while providing a clear fitness advantage in the environment of constant antibody selection encountered in vivo. Thus, sites that can change with little or no impact on relative fitness may have evolved as an immune evasion mechanism to facilitate rapid and early escape from Env-specific antibody.
Acknowledgments

I would like to begin my acknowledgments with the faculty members of the Virology program, including the admissions committee the year that I applied to the program and the faculty members who I interviewed with that year, including Dr. Lee Gehrke, Dr. R. Paul Johnson, and Dr. Welkin Johnson. As a first-generation college graduate, a minority from a disadvantaged background, and a military veteran, each of the members who saw in me the potential to not only be a student within the program, but a colleague within the ranks of other Harvard scientists changed the course of my career and life. I would also like to thank the members of my dissertation advisory committee, Dr. Lee Gehrke, Dr. Karl Munger, and Dr. Miti Kaur for steering me through my graduate studies and providing invaluable guidance on my research.

Next I would like to thank all the members of the Virology program family including all of the other graduate students who I got a chance to meet and interact with during my time in the program and the faculty members who took the time out of their schedules to teach the courses we took in our first and second years. I would especially like to thank my entering class of 2009, who over the course of our time before joining labs became a family to me.

I would like to acknowledge all of the members of the Johnson lab who I’ve gotten to know over the last five years. I especially want to thank Dr. Ben Fofana. He has been a tremendously valuable mentor to me in the lab, who patiently taught me all he knew about antibodies, protein purification, and neutralization assays when I first started in the lab!! But more than that he has become a dear friend to me. We have had conversations on just about every possible topic ranging from politics to religion to raising our families in Boston. On that note, I would also like to thank Alison Hill, who, if I can use a military analogy, went through bootcamp with me in Longwood, before we both got stationed at Camp Welkin Johnson!! I would first like to thank Alison for playing a role in putting my research project on the right track and introducing me to the wonders of next-generation sequencing, the unidirectional flow of sample handling, and those amazing excel sheets!! We have witnessed each other mature as scientists
and in life, even if we didn’t want to admit it. Thank you for the friendship you have shown my family
and me over the years. I know somehow we will see you in Kentucky soon!! I want to thank Kevin for
showing me what passion for science looked like and teaching my kids how to curse! I think you will
always be the only track-running death-metalhead I know. Also I would like to thank Annabel who sat
next to me for just under 4 years and showed my how efficient I could have been in the lab, but was not.

Finally, I would like to thank Welkin first for accepting me into his group and taking on the
challenging role of being a graduate student mentor and for his continuous support through my entire time
in his lab. I want to thank you for allowing me to pursue my line of research in the lab and mentoring me
through the earliest stages of the project forcing me to think critically about my work. Also, thank you
for giving me the flexibility I needed to maintain my responsibilities as a graduate student and my
responsibilities as a father.

I need to thank my family starting with my Mom. Thank you for instilling in me a sense to go
out into the world and pursue something bigger then myself, for teaching me the importance of an
education, and teaching me to believe in myself. You will always be my first inspiration for all that I
have accomplished in my life. I want to thank my brother Louie. We would probably never have
predicted growing up in New Mexico that we would end up together for so many years here in New
England, but here we are! Thank you for producing the artwork for the 2012 program T-shirt that is now
worn by over 50 Harvard virologists!

I would like to thank each one of my children-Ava, Amaya, and Santi. You give my life meaning
and everyday since you were all born have taught me how to be more patient and more loving.

Last, I must thank my wife April. When we first met I was just about ready to leave the military
and all I had to share with you was a dream that I was going to be a doctor one day. It only took me 11
years, but I did it! You have been with me at every step of this journey. We had no idea what we were
getting ourselves into when we came to Boston, but in our time year we have become stronger together in
almost every way. I thank you for doing everything!!! For providing a home to the kids, for pushing me
when I thought that I couldn’t keep up, for keeping me together!
# Table of Contents

Abstract ......................................................................................................................................................... iii

Acknowledgments ........................................................................................................................................... v

List of Figures .................................................................................................................................................. x

List of Tables .................................................................................................................................................. xiii

CHAPTER 1  Introduction-Replication and Persistence of Primate Lentiviruses ........................................ 1

1.1 Primate lentiviruses are complex retroviruses...................................................................................... 2

1.2 The Persistence of PLVs through Immune Evasion and Viral Sequence Diversity ............................. 13

1.3 Structure and Function of PLV Env: Role in viral entry and cellular tropism ................................. 15

1.4 The antibody response to PLV Env....................................................................................................... 19

1.5 The Evolutionary Dynamics of PLVs ................................................................................................... 27

1.6 Scope of dissertation ............................................................................................................................ 39

CHAPTER 2  Distinctive Patterns of Evolution Consistently Associated with SIV\textsubscript{mac251} Env during Early Infection .................................................................................................................. 41

2.1 Acknowledgments ................................................................................................................................. 42

2.2 Abstract .................................................................................................................................................. 43

2.3 Introduction ............................................................................................................................................ 45

2.4 Materials and Methods ......................................................................................................................... 49

2.5 Results .................................................................................................................................................... 58

2.5.1 Antibody responses to SIV Envelope in cohort of SIV-infected rhesus macaques ......................... 58

2.5.2 High-resolution deep sequencing of the SIV env gene during replication in rhesus macaques .... 59

2.5.3 Diversity in stock of uncloned SIV\textsubscript{mac251} swarm captured by Illumina deep sequencing .... 60

2.5.4 Loss of diversity and divergence from the stock reflects a transmission bottleneck during mucosal transmission of uncloned SIV\textsubscript{mac251} .................................................................................................................. 62

2.5.5 Reestablishment of diversity from a homogeneous population during acute infection ...... 64
2.5.6 Effect of antibody escape changes on Env conformation using Env-specific Rh mAbs ................................................................. 82
2.5.7 Detection of ADCC responses in SIVmac251-infected rhesus macaques ................................................................. 87
2.5.8 Individual adaptations do not confer resistance to ADCC activity in SIVmac251-infected rhesus macaques ................................................................. 88
2.6 Discussion................................................................................................................................................................................. 91

Chapter 3 The Relative Viral Replicative Fitness of Env Adaptations Measured Using Fit-Seq .... 102
3.1 Acknowledgements ........................................................................................................................................................................103
3.2 Abstract..................................................................................................................................................................................... 104
3.3 Introduction ................................................................................................................................................................................106
3.4 Materials and Methods ..............................................................................................................................................................110
3.5 Results ................................................................................................................................................................................................ 116
  3.5.1 Viral Fitness of in vivo SIV Env adaptations ................................................................................................................116
  3.5.2 Fit-Seq assay variability .................................................................................................................................................... 120
  3.5.3 Antibody escape adaptations are not associated with a loss of viral fitness .............................................................. 122
  3.5.4 Env V4 loop adaptation has fitness advantage in presence of Env-specific Rh mAb .......................................................... 123
3.6 Discussion................................................................................................................................................................................................ 127

CHAPTER 4 Conclusions........................................................................................................................................................................132
4.1 Understanding viral population dynamics and viral escape at the earliest stages of SIV infection ........................................................................................................133
4.2 The advantage of NGS deep sequencing methods over standard sequencing to understand the dynamics of viral escape. ........................................................................................................133
4.3 P421 represents a key site to mediating antibody escape without associated viral replicative fitness costs ......................................................................................................................................................... 135
4.4 Longitudinal deep sequencing of SIV populations provides the opportunity to study other regions of the SIV genome ........................................................................................................................................................................135
4.5 Implications of virus neutralization missed by standard neutralization assays but detected by Fit-Seq assay ................................................................. 136

4.6 Final Remarks ........................................................................................................ 138

References .................................................................................................................. 140

Appendix .................................................................................................................... 169
List of Figures

Figure 1. The genome organization of two primate lentivirus lineages ................................................. 3

Figure 2. Schematic representation of a mature HIV/SIV virion ............................................................... 5

Figure 3. The lifecycle of primate lentiviruses represented by HIV .......................................................... 9

Figure 4. Summary of the entry process of primate lentiviruses HIV and SIV ........................................... 10

Figure 5. Subunit organization and structure of Env. .................................................................................. 16

Figure 6. Plasma viral RNA load and antibody-based immune responses following repeated low-dose SIVmac251 mucosal challenge ........................................................... 59

Figure 7. Location of two primer set to produce amplicons spanning the env ........................................... 60

Figure 8. Read coverage of SIVmac251 env gene ....................................................................................... 61

Figure 9. Non-synonymous and synonymous diversity present in an uncloned stock of SIVmac251 ..... 62

Figure 10. Diversity of env gene in virus populations at acute SIV infection of (~2-3 weeks) .......... 63

Figure 11. Sites of divergence from the SIVmac251 stock inoculum at two or three weeks post-infection (acute infection) in four rhesus macaques challenged by repeated low dose challenge .......... 64

Figure 12. Loss of viral diversity during transmission of SIVmac251 before emergence of viral diversity during early SIV infection in rhesus macaques .............................................................. 65

Figure 13. Evolution of sequence diversity in SIV Env V1 loop in Mm 10-10, Mm 198-08, Mm 174-08, and Mm 156-10 ......................................................................................................................... 67

Figure 14. Evolution of sequence diversity in SIV Env V1 loop in Mm 10-10 .................................................. 68

Figure 15. Evolution of sequence diversity in SIV Env V1 loop in Mm 198-08 .............................................. 70

Figure 16. Consensus sequences of the Env V1 loop region in Mm 198-08 over early SIV infection .... 71

Figure 17. Evolution of sequence diversity in SIV Env V4 loop in Mm 10-10, Mm 198-08, Mm 174-08, and Mm 156-10 ......................................................................................................................... 72

Figure 18. Evolution of sequence diversity in SIV Env V4 loop ................................................................. 73

Figure 19. Evolution of sequence diversity in SIV Env V4 loop observed in Mm 10-10 ........................... 76

Figure 20. Evolution of sequence diversity in SIV Env V4 loop observed in Mm 198-08 ......................... 77
Figure 21. Evolution of sequence diversity in SIV Env V2 loop.

Figure 22. Evolution of codon 345 at the base of the V3 loop.

Figure 23. Summary of sites of adaptation in the gp120 subunit of SIV Env.

Figure 24. SIVmac gp120 V4 loop is target of Env-specific antibodies leading to substitutions and in-frame deletions.

Figure 25. Env P421Q substitution leads to loss of binding of V4-loop specific mAbs to SIVmac239 gp140 soluble trimers.

Figure 26. Substitutions at V4 loop P421 site lead to a loss of recognition to V4 loop-specific mAbs.

Figure 27. Env V4 P421Q confers neutralization resistance to V4-loop specific mAbs when tested in the context of neutralization sensitive SIVmac316.

Figure 28. SIVmac239 in vivo antibody escape adaptations do not confer resistance to neutralization by sCD4.

Figure 29. End-point plasma from SIVmac251-infected rhesus macaques have detectable ADCC activity against SIVmac239 WT.

Figure 30. SIVmac239 Env neutralization escape mutants do not confer resistance to ADCC.

Figure 31. SIVmac in vivo Env V4 deletions to not confer resistance to ADCC.

Figure 32. β-galactosidase activity of 239 Env WT and mutant infected TZM-bl cells.

Figure 33. Fit-Seq: Deep sequencing based viral fitness assay.

Figure 34. Env V67M mutation increases replication fitness of SIVmac239.

Figure 35. Env V67M mutation increases replication fitness of SIVmac239 at two different inputs.

Figure 36. Plot of reads of WT and mutant Env V67M produced from two separate Illumina runs.

Figure 37. SIV Env in vivo adaptations do not incur a replicative fitness cost.

Figure 38 Env V4-loop-specific mAb C26 binds to recombinant SIVmac239 gp140 but does not neutralize SIVmac239 by standard neutralization assay.

Figure 39. Env P421Q provides a replicative fitness advantage in the presence of V4-loop specific mAb C26.
Figure 40. Adaptations at P421 in V4 loop provide a mechanism for antibody escape.
List of Tables

Table 1. Cohort of four Indian-origin rhesus macaques challenged by repeated low-dose of uncloned SIVmac251 as unvaccinated control animals. .............................................................. 58

Table 2. Relative fitness of SIV$_{mac239}$ Env V67M ........................................................................................................ 119

Table 3. Relative Fitness of SIV$_{mac239}$ Env V67M at two ratios of WT:mutant virus inputs ................. 120

Table 4. Summary information from two Illumina sequence runs using the same library ..................... 121

Table 5. Relative fitness of SIV Env adaptations .............................................................................................. 123

Table 6. Relative fitness of Env P421Q in presence and absence of V4-loop specific mAb. ................. 125
CHAPTER 1

Introduction - Replication and Persistence of Primate Lentiviruses
1.1 Primate lentiviruses are complex retroviruses

Primate lentiviruses (PLVs), including human immunodeficiency virus type 1 (HIV-1), HIV-2, and the group of simian immunodeficiency viruses (SIVs), belong to the larger family of Retroviridae which all possess single-stranded positive-sense RNA genomes. The two distinguishing features common to all retroviruses are the reverse transcription of their single stranded RNA (ssRNA) genomes into double stranded DNA (dsDNA) and the integration of the viral genome into the genome of the infected target cell. The integrated viral genome is known as the provirus and it is maintained within the host genome and passed on to any progeny cells. This replication strategy is unique in the viral world and underlies the persistent infections PLVs establish within their hosts. Lenti is latin for “slow” and signifies the persistent infections established by all lentiviruses in mammalian hosts including sheep, horses, cows, felines, and primates [1]. Although, all lentiviruses share tropism for macrophages, only primate lentiviruses use CD4 as a primary receptor, in addition to chemokine coreceptors, thus gaining tropism for CD4+ T cells [1]. The targeting of CD4+ T cells holds important implications into disease progression observed in susceptible hosts, specifically under the context of alternative non-natural hosts through cross-species transmission.

Primate lentiviral genomes and their gene products

The primate lentiviral genome is a single-stranded RNA genome of about 10 kilobases, encodes nine viral gene products, and contains repetitive sequence elements (R regions) near the 5’ and 3’ ends (Figure 1). Internally flanking the R regions are elements unique to the 5’ end and 3’ end of the genome referred to as the U5 and U3 regions. Both R regions, the U5 region, and the U3 region are critical to mediating reverse transcription and result in two identical long terminal repeat regions at each end of the newly reverse-transcribed DNA genome. Three of the viral genes, gag, pro, and env are common to all retroviruses and are sufficient for viral replication and virion formation. They are all expressed as polyprotein precursor proteins which are then cleaved by either viral or host cell proteases to produce the mature virus particle-associated proteins [2]. gag (group-specific antigen) codes for the Gag polyprotein, which is processed to produce matrix, capsid, nucleocapsid, and p2, which collectively make up the core
structural proteins of the virion. pol-pro (polymerase) codes for the three viral enzymes protease (PR), reverse transcriptase (RT), and integrase (IN). env (envelope) codes for the envelope glycoprotein precursor, gp160, which is processed to produce the gp120 surface subunit and the gp41 transmembrane subunit. Additionally, PLVs contain six accessory genes, which primarily function to increase viral pathogenicity. tat, rev, nef, vpr, and vif are found in all PLVs, but two gene products, vpu and vpx are not shared by all PLVs. HIV-1 and its ancestors encode vpu whereas SIVsm, SIVmac, and HIV-2 all encode vpx (Figure 1).

**Phylogenetic relatedness of HIV-1 and HIV-2**

Phylogenetic analyses of HIV-1 based on various regions of the genome reveal several different groups including a M (main), O (outlier), N (non-M, non-O), and P group [3]. The M group makes up 95% of the strains in the HIV-1 pandemic and is further classified into clades or subclades that are mostly distributed geographically [2]. The HIV-1 M group clades are A, B, C, D, F, G, H, J, and K and additionally include 15 circulating recombinant forms (CRFs) which are recombinants of two or more viruses from different clades [2, 3]. Of note, clade B is prevalent in North America and Western Europe and clade C is prevalent in sub-Saharan Africa, which contains the highest prevalence rates of the
pandemic. HIV-2 is also categorized into several lineages, but its distribution is much more limited to Africa reflective of a less severe pathogenesis and progression to AIDS relative to HIV-1 [3]. The large degree of heterogeneity of HIV and CRFs of HIV-1 reflect the high rate of recombination and error-prone viral replication inherent in PLVs.

**PLV virion structure and make-up**

Virions of PLVs range in size from 100 to 120 μm and are enveloped by a host-cell derived lipid membrane [2]. The envelope glycoprotein (Env) is embedded in the lipid membrane as a heterodimeric trimer complex (Figure 2). Mature virions are lined by matrix (MA) underneath the lipid membrane and contain a higher-order capsid core that surrounds two copies of the ssRNA genome that is encapsidated by nucleocapsid (NC) and viral enzymes, most importantly reverse transcriptase (RT) and integrase (IN). The capsid core is made up of both hexamers and pentamers of the capsid (CA) monomer. The 12 CA pentamers, relative to hexamers, produces the curvature of the fullerene shape of the capsid core. Encapsidation of the ssRNA genome and viral enzymes by the capsid core largely carries out two important functions. First, it provides a barrier to host innate immune factors that would otherwise recognize the viral RNA upon entry into the target cell. Secondly, it provides a microenvironment that keeps the ssRNA genome in close proximity to RT and facilitates reverse transcription.

**HIV is the causative agent of acquired immunodeficiency disease syndrome (AIDS)**

HIV infection in humans occurs through exchange of bodily fluids and HIV is a common sexually transmitted disease, however, transmission also occurs through sharing of needles between HIV-positive individuals with others. The primary target cell of HIV is CD4+ T lymphocytes, otherwise known as T-helper cells. Upon successful transmission HIV replicates to high titers within 1-2 weeks and typically induces flu-like symptoms in the infected host. Titers at acute peak of viral replication may reach an average of $10^7$ viral copies per milliliter (ml) of blood [4]. Subsequent to acute peak of viral replication viral loads decrease with the onset of innate and adaptive immune responses, but the virus is not cleared and viral loads reach a set point [5]. During this period of infection HIV replicates continuously establishing a persistent infection that may last for months to years in the infected host and
Figure 2. Schematic representation of a mature HIV/SIV virion.
The Env trimer is displayed at the surface of the lipid membrane of the virion. The TM region of gp41 subunit anchors the Env complex in the membrane. The gp120 subunit and the ectodomain of the gp41 subunit are surface exposed. MA lines the internal surface of the membrane. CA encloses the two copies of the single-stranded genomic RNA (gRNA) encapsidated by NC. Similarly, RT and IN are found in the conical CA core. Viral proteins and virion are not to scale.

is defined as clinical latency [6, 7]. During this period of clinical latency viral loads and CD4⁺ T cell counts remain stable, however, the rate of HIV virion production is near 10⁸ to 10⁹ particles per day and the rate of CD4⁺ T cell turnover reaches up to 2 x 10⁹ cells per day. Thus, this steady state period of infection during clinical latency is maintained through high rates of virion production, CD4⁺ T cell targeting and killing, and generation of new uninfected CD4⁺ T cell targets [8]. In the absence of any antiretroviral treatment most individuals eventually progress to AIDS. Late stage disease is characterized by the reduction of CD4⁺ T cells to a low threshold, typically <200 cells/ml and an increase in viral loads. The loss of CD4⁺ T cells leads to the development of immunodeficiency that characterizes late-stage
AIDS leading to the development of rare, opportunistic infections previously only observed in immunocompromised patients including fungal infections such as *Pneumocystis* pneumonia and candidiasis [9]. Rare opportunistic cancers now associated with the development of AIDS include Karposi’s sarcoma and Burkitt’s lymphoma [10]. Prior to the development of antiviral drugs against HIV the period between acute peak of viral replication and the development of late-stage AIDS were dependent upon the dynamics of the infecting stain of virus and the immune response of the infected host. Several classes of antiretroviral drugs, which target the different stages in the retroviral lifecycle including entry, reverse transcription, integration, transcription, assembly, and protease processing have been developed [11]. When taken in combination antiretroviral drug therapy (ART) is effective in reducing viral loads below the threshold of detection and more recently different combinations of 3 to 4 antiretroviral drugs have been combined into single pills [12, 13], transitioning HIV-1 infection into a chronic disease that can be regulated through daily adherence to multiple drug regimens.

Despite the effectiveness of ART, it is not successful in eradicating the virus in the host. Any interruption in ART leads to a rapid return of detectable viral loads, even after long-term treatment [14]. Several studies discovered viral reservoirs present in cells with latent integrated provirus as the source of the viral load rebound [15-17]. Current research is aimed at defining strategies to activate these long-lived cell populations and then eradicate them [18]. However, the lack of any foreseeable strategy highlights the need for a safe and effective prophylactic vaccine against HIV.

**The emergence and current state of HIV-1 induced AIDS as a global pandemic**

The current HIV/AIDS global pandemic emerged suddenly and without notice around 1981 when cases of individuals succumbing to immunodeficiency-related deaths occurred in Los Angeles and other large cities [19-21]. Initial cases were more prevalent in high-risk transmission groups including injection drug users and homosexual men, and a reduction in CD4+ T cells was a common observation in initial cases, explaining the acquired immunodeficiency. The observation that many partners of individuals with disease from high risk groups, hemophiliacs, and individuals who received blood transfusions suggested a causative agent was responsible for the disease and HIV-1 was eventually identified as the causative
agent responsible for AIDS [22, 23]. The AIDS epidemics in the United States and other Western
European nations were eventually realized to be a part of a larger global pandemic of HIV-1/AIDS with
the highest prevalence rates occurring in nations of sub-Saharan Africa [3, 24, 25]. Since the beginning
of this modern pandemic over 78 million people have been infected with HIV, over 39 million people
have died of AIDS-related deaths leaving over 36 million people living with HIV [24, 25]. However, we
have turned a corner in the epidemic with areas of the highest prevalence rates, notably sub-Saharan
Africa and the Asian Pacific, reporting dramatically reduced numbers of newly infected individuals
resulting in a 35% global reduction of new infections since 2000 and 42% reduction in AIDS-related
deaths since a peak in 2004 [25]. Much of this progress is related to access to antiretroviral treatment,
however, 22 million individuals still need access to treatment and repeated failures to produce a protective
vaccine indicate HIV/AIDS will continue to be a global public health threat for some time. Despite these
hurdles the UNAIDS has declared a global priority to end the AIDS pandemic by 2030 by increased
treatment rates, decreased disease spread, and hopefully a therapeutic vaccine.

The origins of HIV-1 and HIV-2

The identification of HIV-1 as the causative agent of AIDS prompted investigation into the
sudden emergence of HIV in humans and its spread through the population. The identification of two
additional primate lentiviruses provided the first clues as to the origins of HIV. First was the discovery of
a primate lentivirus that induced an AIDS-like disease state in rhesus macaques housed in US primate
centers, named simian immunodeficiency virus (SIV) based on its pathogenesis and relatedness to HIV
[26]. Second was the discovery of another human lentivirus with a similar morphology to HIV-1, but
distinct antigenic properties, found in AIDS patients in Western Africa [27]. SIV was soon found in
many species of African primates, including African green monkeys, sooty mangabey monkeys,
chimpanzees, and gorillas of sub-Saharan Africa [28-30]. SIV was eventually identified in over 40
species of African primates and strikingly did not appear to induce any pathologies or AIDS-like
symptoms in the natural host despite the detection of actively replicating virus at high viral loads [3].

7
Phylogenetic analysis find SIV strains of a given species cluster together revealing that most SIVs are adapted to the host species they infect, however, numerous studies have found evidence of cross-species transmission events among primate species and in some instances the emergence of mosaic SIVs through super-infection and recombination [31-37]. Naturally these host-species jumps have resulted in differences in prevalence and pathogenesis in African primate hosts with longer periods of virus-host co-evolution likely resulting in adaptation leading to non-pathogenic state in the host [38]. The ability of PLVs to induce disease, therefore, likely reflects a more recent cross-species jump such as HIV-1 and HIV-2 infections of humans and SIVmac infection of rhesus macaques.

SIVs from chimpanzees are found to be the most closely related to HIV-1 amongst SIVs from African primate species [3] and multiple reports confirm HIV-1 is a cross-species transmission from the chimpanzee sub-species *Pan troglodytes troglodytes* which harbor SIVcpzPtt [39, 40]. The jump of SIVcpzPtt from chimpanzees into human is estimated to have occurred around 1920 leading to the emergence of HIV-1 group M, the main group of the pandemic [41]. Phylogenetic analysis reveals four other groups of HIV-1, indicating HIV-1 has emerged in humans in at least four separate events leading to the different groups. HIV-2 and SIVmac are both more similar to SIVs of sooty mangabey monkeys (SIVsm) [42], thus HIV-2 represents multiple jumps of SIVsm into humans while SIVmac was the result of accidental and intentional exposure of Asian-origin macaques to SIV from sooty mangabey monkeys [43, 44].

**The virus lifecycle of SIV and HIV**

*Env-mediated viral entry.* Starting from a cell-free infectious mature virion the first step in the virus life cycle is host cell attachment and fusion of the membrane of the virion with the membrane of the target cell (Figure 3). PLV particles are enveloped by a lipid bilayer membrane that is derived from the infected cell, and is embedded with mature envelope glycoprotein trimers. Entry requires cellular attachment through cellular receptor binding and subsequently fusion of the viral membrane with the target cell membrane (Figure 4). Similar to other class I fusion proteins, PLV Env mediates these two events with two subunits [45]. The gp120 surface subunit mediates cellular receptor binding and contains
the primary and coreceptor binding sites. The gp41 transmembrane subunit has a N-terminal ectodomain, a transmembrane domain, and a C-terminal cytoplasmic tail. The fusion peptide, which mediates fusion, is contained in the N-terminal ectodomain of gp41.

Figure 3. The lifecycle of primate lentiviruses represented by HIV.
Starting with the cell-free virion, the basic steps in the lifecycle of primate lentiviruses begin with Env mediating cellular receptor attachment and fusion of the virion with the host cell. RT reverse transcribes the ssRNA genome into dsDNA, which is then translocated across the nuclear membrane. Once in the nucleus IN mediates integration of the viral genome into the chromosomal DNA. Tat and Rev function to promote viral gene transcription and transport of unspliced or incompletely spliced transcripts into the cytoplasm for gene expression, respectively. Assembly occurs at the lipid membrane with Gag providing the structural framework of the budding virion and interacts with Env trimer complexes to promote Env uptake into the newly budding virion. The late domain of P6 mediates the final budding step of the membrane using host cell machinery. In the immature nascent virion PR cleaves Gag and Pol into final gene products associated with the maturated virion, leading to formation of the conical CA core that surrounds the two copies of the ssRNA genome encapsidated by NC. Each step is described in detail in the text of this section. (Adapted from Rambaut et al., 2004, with permission [46]).

The primary receptor common to all PLVs is the CD4 molecule, which is present on various cells of the immune system, but is highly expressed on CD4+ T lymphocytes [47-49]. In the first step of the entry process, CD4 binds to the conserved CD4 binding site on gp120, which triggers conformational rearrangements of the gp120 subunit that result in the formation and exposure of the coreceptor binding
site and primes gp120 to engage coreceptor molecules [50-52]. Coreceptor molecules for PLVs are all multimembrane spanning chemokine receptors, but the most commonly used coreceptors are CCR5 and CXCR4 [53-56]. Binding to coreceptor molecules results in further conformational rearrangements that induce gp120 dissociation and expose the gp41 ectodomain [45]. The exposed state of gp41 triggers the release of the fusion peptide into the target cell membrane forming the prehairpin intermediate and bringing the viral membrane into close proximity with the target cell membrane [57-60]. The gp41 ectodomain contains two alpha-helical heptad repeat regions downstream of the N-terminal fusion peptide region [57, 60]. The fusion peptide brings the N-terminal heptad repeat region (HR1) into close proximity with the target cell membrane, while the C-terminal HR2 is located near the TM domain embedded in the viral membrane.

Figure 4. Summary of the entry process of primate lentiviruses HIV and SIV.
1. The Env trimer complex consists of gp120:gp41 heterodimers embedded in the viral membrane by the TM region of gp41. 2. The first step in the entry process is gp120 binding to the primary receptor CD4 attaching the virion to the target cell. 3. gp120 interaction with CD4 induces conformational changes that expose the coreceptor binding site allowing for binding with the coreceptor, typically CCR5 or CXCR4. 4. Interaction with coreceptor results in further conformational rearrangements exposing the fusion peptide, which is released into the target cell membrane, leading to the formation of the six helix bundle, and finally fusion of the viral membrane with the host cell membrane. (Adapted from Wilen, C.B., Tilton, J.C., and Doms R.W. with permission [61]).

HR1 collapses at a hinge between the two HR regions onto HR2 resulting in the formation of a six-helix bundle made up of HR1 and HR2 from each gp41 molecule in the trimer [45]. The meta-stable state of the six-helix bundle triggers the mixing of the viral and cellular membranes resulting in the opening of a fusion pore that allows the capsid core to enter into the cytoplasm and begin the next step in the virus lifecycle.
Reverse transcription of the RNA genome into dsDNA and integration into the host genome. The post-entry step of the PLV life cycle requires uncoating of the capsid core, reverse transcription of the RNA genome into a dsDNA template, and delivery of that template into the nucleus for integration into the host genome. All of the necessary enzymes to carry out this process are present in the entering conical capsid core. The RNA genome of PLVs consists of two single stranded linear copies. Reverse transcriptase initiates transcription using a lysine tRNA which is complementary to the primary binding site (PBS) found near the 5’ end of the genome [62]. RT transcribes this short stretch of RNA to produce minus-strand DNA through to the end of the genome. The RNase H domain of RT degrades RNA in the newly synthesized DNA:RNA duplex leading to minus strand strong-stop DNA (-sssDNA). The first of two strand transfers occurs when the PBS stretch of –sssDNA anneals to the complimentary PBS present near the 3’ end of the genome. DNA synthesis of the remaining RNA genome occurs through to the PBS. RNase H degrades the RNA template except for a small stretch of RNA known as the polypurine tract (PPT). The PPT is resistant to degradation and effectively becomes a primer to synthesize positive strand DNA using the minus-strand DNA as a template. DNA is synthesized through the tRNA primer producing plus-strand strong-stop DNA (+sssDNA). The small stretch of tRNA is degraded by RNase H exposing the newly produced +sssDNA leading to the second strand transfer at the complimentary PBS sequence present in both newly synthesized DNA strands. This final transfer allows for both strands to be completely transcribed leading to a linear blunt ended dsDNA molecule that is complete for integration into the host genome.

The linear DNA molecule enters into a pre-integration complex that contains the viral enzyme, integrase (IN), matrix (MA) and Vpr present in the entering capsid core [63, 64]. The PIC must access the host genomic DNA either by translocation across the nuclear membrane or breakdown of the nuclear membrane during the mitosis step of the cell cycle. PLVs are able to infect non-dividing cells therefore the PIC is translocated across the nuclear membrane, in contrast to simple retroviruses which must wait for breakdown of the nuclear membrane limiting their infection to dividing cells.
The mechanism of integration involves two reactions, which are both catalyzed by integrase and occur through the same transesterification reaction which involves a nucleophilic attack of a phosphodiester bond by a hydroxyl group [65]. 3’ end processing results in the removal of a dinucleotide at each 3’ end of the linear viral DNA molecule producing a two base pair overhang on the 5’ end and newly exposed hydroxyl group on the 3’ end. In the DNA joining or strand transfer step the 3’ hydroxyl groups on each viral end attack phosphodiester bonds in opposite strands of the target cellular DNA effectively joining the viral DNA strands into the host DNA. The target sites in the DNA are typically separated by 5 base pairs for PLV integration, but vary with other retroviruses [66]. Completion of insertion occurs by subsequent removal of overhangs, DNA synthesis to fill in the gaps, and a ligation to fill in the nick on each strand of DNA. The newly synthesized DNA to fill in the missing gap produces 5 base-pair repeats of cellular DNA flanking the newly inserted viral DNA that is characteristic of PLV insertions.

**Viral transcription and translation from the integrated provirus.** The long terminal repeats (LTRs) function as promoters for the transcription of spliced and unspliced templates in addition to guiding insertion of the proviral DNA into the host cell chromosomal DNA. Promoter activity is primarily driven from the 5’ LTR [2]. A second region critical to viral gene expression is the cis-acting rev response element (RRE), which is located in the env open reading frame. tat and rev are transcribed early and both viral proteins are critical to the expression of viral gene products . Tat is critical to regulation of viral transcription and it’s activity is required for the formation of progeny virions [2]. Tat functions primarily to bind to transcription complexes increasing the steady-state levels of viral RNA and the number of longer RNA species synthesized during transcription. Rev functions primarily to help transport unspliced or incompletely spliced viral transcripts from the nucleus into the cytoplasm to ensure their expression into viral gene products or incorporation into virions during assembly [2]. Rev mediates transport by binding to the RRE element. Thus, the RRE is found in every unspliced and incompletely spliced transcript that is to be successfully transported to the cytoplasm.
**Viral assembly and release.** Viral assembly and release is largely mediated by the Gag-encoded structural proteins MA, CA, NC, and p6. MA functions to provide the structural framework of the nascent virion, and directs the Gag precursor protein PR55\(^{gag}\) to the plasma membrane of the cell through a membrane binding domain and a myristic acid moiety covalently bonded to the N-terminal glycine of MA [2]. A small region of MA (between residues 84 and 88 in HIV-1) directs MA to the plasma membrane and mutations within this region result in a redirection of MA-mediated PR55\(^{gag}\) targeting from the cell surface membrane to intercellular membranes [67, 68]. The targeting of MA to the cell-surface lipid membrane also facilitates the incorporation of Env trimers into budding virions through direct interactions between MA and the gp41 cytoplasmic tail of Env [69-74]. The p6 protein is necessary for the efficient budding of the nascent virion [75]. A highly conserved motif within p6 of PLVs that consists of Pro-Thr-Ala-Pro (PTAP) regulates budding by mediating direct interactions with the Tsg101 component of the ESCRT-I complex [76-78]. PTAP is also commonly referred to as the late domain, which is in reference to the requirement for PTAP at a late stage of the budding process. Mutation of the late domain disrupts budding and release of nascent immature virions [75].

Soon after budding the viral protease (PR) enzyme targets the Gag-Pol precursor proteins for proteolytic cleavage to produce MA, CA, NC and P6 from Gag and RT, IN, and PR from Pol leading to the formation of the mature viral particle [2]. Cleavage of Gag allows CA monomers to come together to form the higher-order capsid core that surrounds the two copies of NC-bound single-stranded genomic RNA, RT, and IN. MA remains attached to the lipid membrane of the virion. At this stage the mature virion particle is able to infect a new target cell.

### 1.2 The Persistence of PLVs through Immune Evasion and Viral Sequence Diversity

All viruses are obligate parasites that require a continuous chain of transmission from host to host. Viral strategies of survival range from acutely infecting viruses with rapid transmission between hosts to persistently infecting viruses with lower transmission rates and longer incubation periods [79]. Acutely infecting viruses typically replicate rapidly before the onset of adaptive immune forces and typically use efficient transmission routes, such as airborne transmission, to successfully infect the next
host. In contrast, transmission by persistent viruses typically requires the capacity to deal with host immune responses. Viruses that persist through latency are reliant on periods in which the genome is not expressed, thereby remaining silent to the host immune response. Transmission occurs when replication becomes active but viral replication is controlled fairly rapidly. Thus periods of latency are required to maintain transmission. PLVs maintain persistence through continuous, uninterrupted replication, a rare strategy among viruses, which differs from persistence through latency. Persistence through continuous replication provides a ready substrate for the host immune response; therefore, PLVs have evolved sophisticated strategies of immune evasion to overcome adaptive immune responses [79, 80].

The envelope glycoprotein is the primary target of the B cell immune response mediated by antibodies and is a profound example of the co-evolution between virus and host. As PLVs are enveloped viruses and must fuse with the target cell, Env has evolved under dual selective forces. First, Env must maintain structural conservation that allows for proper synthesis, processing leading to expression on the cell surface, and uptake into the membrane of the budding virion to mediate fusion with the target cell. Second, Env must also maintain this function under constant antibody surveillance. Antibodies can bind and block interaction with host cell receptors resulting in neutralization of the virus. Antibodies that bind Env can also recruit other immune effector functions such as complement or NK cells leading to NK-cell mediated killing of Env-expressing infected cells. Thus, Env must evade antibodies with a diverse array of effector functions. Env has evolved several strategies to reduce antibody targeting of conserved receptor binding sites (primary and coreceptor) present in the gp120 surface subunit of Env. In addition to immune-selected features of the Env trimer, immune evasion through viral sequence variation, leading to viral escape, remains the paramount immune evasion strategy that overcomes host antibody responses and all strategies to produce protective adaptive immune responses by vaccination. The remaining sections will cover what is known about the antibody response to Env, immune evasion strategies, the adaptability of Env through sequence variation and the scope of this dissertation to understand the evolution of viral escape during acute/early SIV infection.
1.3 Structure and Function of PLV Env: Role in viral entry and cellular tropism

Synthesis and processing of Env

The envelope glycoprotein (Env) is expressed from the env gene as a 160 kilodalton (gp160) protein precursor of approximately 845-870 amino acids and trafficked to the endoplasmic reticulum (ER) by a N-terminal signal peptide [81]. Once in the ER Env is modified through numerous N-linked, high mannose glycans that are necessary for proper folding and oligomerizes into trimer complexes through interactions mediated by the N-terminal ectodomain of gp41 [82]. The gp160 Env trimers are transported to the Golgi apparatus whereupon each gp160 molecule is cleaved by cellular furin-like proteases into two subunits, the gp120 surface subunit (SU) and the gp41 transmembrane subunit (TM). Each heterodimer remains non-covalently linked through interactions between the N-terminal and C-terminal regions of gp120 and the ectodomain region of gp41. Thus, the final mature form of Env is made up of a trimer complex of gp120:gp41 heterodimers (Figure 5). Env trimers are finally trafficked to the plasma membrane and during viral assembly a subset of Env trimer complexes are incorporated into viral particles resulting in a surprisingly small number of 7-14 mature intact Env trimers found on the virion particle [83-86].

Structural features of the Env trimer complex

gp120 consists of five conserved regions (C regions), which largely mediate primary and coreceptor interactions, interspersed between five variable regions which play a pivotal role in antibody immune evasion and protection of conserved regions of gp120. Four of the five variable regions exist as loops due to the presence of intermolecular disulfide bonds between highly conserved cysteine residues that flank each of the variable regions. A crystal structure of the HIV-1 gp120 monomer found the conserved regions exist as an inner domain, an outer domain, and a bridging sheet made up of four anti-parallel β-strands that connect the two domains [87]. The domain names reflect the orientation of the domains relative to the trimer complex. Thus the inner domain exists along the internal axis of the trimer, where it is mostly buried. In contrast the outer domain is mostly exposed on the surface of gp120 and contains several sites of N-linked glycosylation.
Figure 5. Subunit organization and structure of Env.
A. A schematic of Env representing the gp120 surface (SU) subunit and the gp41 transmembrane (TM) subunit. Conserved regions (C regions) and variable regions (V regions) of gp120 are interspersed. The 23 sites of N-linked glycosylation in gp120 and the 3 sites of N-linked glycosylation in gp41 are marked by tree structures above the schematic. B. The crystal structure of a native-cleaved HIV-1 Env gp140 recombinant protein (BG505 SOSIP.664) represents our most current understanding of the full Env trimer complex. The three gp120 subunits of the Env trimer are represented in red and the three gp41 subunits are represented in blue. Dotted lines within the structure are used to represent variable regions that are unresolved in the structure. Of importance is the presence of the V1/V2 domains near the apex of the trimer (facing outwards towards the incoming target cell), the presence of other variable loop regions lining the external surface providing coverage to conserved receptor binding regions, and the level of inter-subunit interaction between gp120 and gp41. (Adapted from a recent review of HIV antibody responses in Burton, D.R. and Mascola, J.R., with permission [88]).

Interestingly, both domains and the bridging sheet contribute to interactions with CD4 [87]. The CD4 binding site is recessed into a pocket of gp120 and contacts with CD4 span an area over 800 Å across the gp120 surface [87]. Critical interactions with CD4 mapped to a deep cavity within the pocket of gp120. The opening of this cavity is filled by CD4 residue phenylalanine 43 and many gp120 residues found to be critical for CD4 interaction by mutational analysis mapped to this deep pocket of gp120, including a highly conserved GGD(P/L)E motif. gp120 interactions with CD4 lead to conformational rearrangements leading to the formation of the coreceptor binding site [89]. Similar to the CD4 binding site, the coreceptor binding site spans several conserved regions of gp120 including the bridging sheet, the base of the V1/V2, and sequences in the C4 region. The V3 loop makes direct coreceptor interactions,
thus the V3 loop is critical to coreceptor usage and cellular tropism [90]. Interestingly, sulfated tyrosines located within the N-terminus of CCR5 were found to be critical for interactions with gp120 [91, 92], leading to a model of interaction in which the V3 loop binds to the second extracellular loop of CCR5 while the N-terminal region, containing sulfated tyrosines, interacts with the coreceptor binding site [93].

CCR5 is the primary coreceptor molecule used by HIV and SIV, however, it has been estimated that over 40% of HIV-1 infected individuals develop variants that utilize CXCR4 [94]. This dichotomy has resulted in the description of HIV and SIV variants as either CCR5-tropic (R5-tropic) or CXCR4-tropic (X4-tropic) variants. The evolution of X4-tropic variants is correlated with increased disease progression, although not all HIV-infected individuals who progress to disease develop X4 tropic variants [95], thus it is possible that X4-tropic variants may be a consequence, rather then a cause, of disease progression. Of note, most transmitted variants of HIV and SIV are R5-tropic suggesting a preference for R5-tropic variants exists during transmission [96]. Interestingly, individuals who are homozygous for deletion in CCR5, resulting in a defective ccr5 allele, are highly resistant to infection, and individuals who are heterozygous show delayed disease progression [97-101]. Of importance, the single case of an individual cured of HIV-1 occurred through a bone-marrow transplant from an individual who was homozygous for the ccr5 Δ32 gene [102, 103]. In addition to CCR5 and CXCR4, several other receptors have been identified for HIV and SIV [104-108].

Adaptations in Env can change cellular tropism. The most relevant change in tropism, in regards to viral pathogenesis, involves adaptations leading to macrophage tropism, which has significant implications in the development of neurological disorders. Several SIVmac strains have been isolated from CNS and brain tissues that replicate well in macrophages in vitro [109]. SIVmac316 was isolated from an SIVmac239-infected animal during late stage disease, found to be macrophage tropic [109], and highly neutralization sensitive [110, 111]. Interestingly, SIVmac316 had 9 changes relative to SIVmac239, with one change in gp41 resulting in a premature stop codon. Of the other eight substitutions only 3 were critical for macrophage tropism [109]. More recently, two Asn residues at sites of N-linked glycosylation were found to be critical for SIVmac tropism for macrophages [112, 113]. Removal of NLGS at N173 and
N481 by mutation of Asn to Gln resulted in increased macrophage tropism, but also increased neutralization resistance. N173 is located in V2 near the apex of the Env trimer and N481 is located in C5 near the CD4 binding site, thus it is suggested that removal of the NLGS at positions 173 and 481 may allow for conformational changes that allow for CD4-independent coreceptor interactions leading to enhanced fusogenicity. Of importance, these adaptations were identified in a SIVmac251 variant isolated from PBMCs during early infection, evidence of the presence of macrophage tropic variants in the viral population as early as two weeks into infection [112].

Comparison of cryo-electron microscopy (cryo-EM) structures of the Env trimer prior to and after binding to CD4 revealed two distinct conformational states [114, 115]. Prior to CD4 binding the trimer is in a “closed” conformation with the base of the V1/V2 loops positioned near the center of the trimer and the primary and coreceptor binding sites covered. However, CD4-bound conformations of the trimer show conformational changes moving the V1/V2 and V3 loops away from the axis of the trimer resulting in an open conformation predicted to allow for coreceptor engagement.

Cryo-EM studies with CD4-independent strains of SIV and recombinant, soluble SIV gp140 trimers found the trimer complex exists in an open conformation prior to CD4 binding [115, 116]. These studies were in line with a model that the structural rearrangements of CD4 binding provided access to the coreceptor binding site. Strains that naturally exist in the open confirmation bypass the need for CD4 engagement, but are rendered more neutralization sensitive.

The gp41 transmembrane subunit is made up of the N-terminal ectodomain, a transmembrane region responsible for anchoring the gp120:gp41 heterodimer in the lipid membrane, and a C-terminal cytoplasmic tail. The ectodomain mediates fusion of the viral membrane with the target cell membrane. It interacts with the gp120 subunit over several covalent interactions that inhibit the dissociation of gp120. It contains the N-terminal fusion peptide which is kept in a restricted conformation until gp120 interaction with CD4 induce conformational changes that lead to the release of the fusion peptide into the target cell membrane. The ectodomain also contains two heptad repeat regions that form coiled-coil
helices [57, 60]. Heptad repeat region 1 (HR1) and HR2 are critical to the conformational rearrangements that directly fuse the viral and cellular membranes, as described in the previous sections.

The gp120 subunit and gp41 subunits maintain covalent interactions after cleavage producing the heterodimer that multimerizes into a trimer during processing in the ER through interactions between the gp41 ectodomains. This heterodimer of trimers forms the mature PLV envelope glycoprotein commonly referred to as the viral spike.

1.4 The antibody response to PLV Env

Early studies of antibody responses focused on the neutralization sensitivity of clonal isolates of HIV-1 and large panels of Env-pseudotyped viruses that contained primary env sequences cloned from HIV-1 infected patients including subtypes B and C [117, 118]. Neutralizing antibody responses measured with plasma or sera develop and mature over time, however, samples typically only neutralized host matched (autologous) virus, but had minimal measurable effect on heterologous viruses [119]. These initial reports were conducted with patients infected with HIV-1 of subtype B, however, similar strain-specific neutralizing antibody responses were observed in patients infected with subtype C providing strong evidence that HIV-1 Env possessed an inherent resistance to neutralization [120, 121]. Studies with larger samples sizes identified broadly neutralizing antibody responses with cross-neutralizing activity, however, these responses were relatively rare and in all cases viral escape was evident through testing of longitudinally-isolated primary Env clones that gained increased resistance to contemporaneous plasma or plasma from earlier time-points of infection [122-124].

Mapping the antibody specificities of these broadly neutralizing antibodies identified the CD4 binding site and the coreceptor-binding site as predominant targets, but targets were also found outside the CD4 binding site [125-127]. One study of early HIV-1 subtype C responses found the target of neutralizing antibody plasma samples from HIV-1 clade C infected patients included all five variable regions, but all four infected patients commonly responded to the C3-V4 region [128]. The estimated time for the earliest cross neutralizing antibody response to develop was ~1 year using plasma from patients infected with clade B isolates tested against primary strains of clades A, B, and C and again
responses were found to target the CD4 binding site and a glycan-dependant epitope at the base of the V3 loop [129, 130].

The isolation and characterization of broadly neutralizing antibodies to HIV-1

More recent studies have focused on the isolation and characterization of monoclonal antibodies (mAbs) responsible for the potency and breadth of neutralization found in these rare plasma samples. The first group of broadly neutralizing antibodies (bNAb) discovered targeted epitopes in the gp120 subunit and the gp41 subunit. bNAb 2F5 and 4E10 both recognized epitopes in the sequence conserved membrane proximal external region (MPER) of gp41 [131, 132]. b12 was isolated through a phage display library, and was one of the first bNAb found to recognize the CD4 binding site [133]. 2G12 was the first mAb that targeted highly conserved N-linked glycans at the base of the V3 loop as part of its epitope [134, 135]. The presence or absence of an NLGS at position 295 dictated whether strains of HIV-1 were sensitive or resistant to 2G12 [136-138]. These mAbs provided insight into the antigenic targets available to host antibodies despite multiple structural features that served to reduce antibody targeting present on the Env trimer [139].

More recently, sophisticated cell sorting methods to isolate B cells that produce HIV-1 Env specific antibodies and engineered forms of soluble Env trimers identified several mAbs with even higher potency and breadth then the first generation of bNAb [140-143]. VRC01 targeted the CD4 binding site, and was one of the first isolated by this new technique [141]. The epitope of VRC01 does not completely overlap with the CD4 binding site or the b12 epitope, however, VRC01 could neutralize over 90% of strains tested, and at a relatively low concentration reduced virus infectivity by 50% (IC_{50}). Several other CD4 binding site antibodies have been isolated providing strong evidence that the CD4 binding site is an effective, if not the most effective, target for neutralization [142, 144]. The rare frequency and the time needed for these antibodies to mature in HIV-1 infected hosts highlights the difficulty in attaining specificity for the CD4 binding site.

PG9 and PG16 are bNAbs that target glycans present at positions 156 and 160 in the V1/V2 loop of HIV-1 [140]. Interestingly, the epitope for PG9 and PG16 is recognized within the context of soluble
trimers or surface-expressed Env, but not monomeric gp120, suggesting recognition of an epitope present only on the quaternary structure of the Env trimer. Crystal structures of PG9 and PG16 revealed a CDR H3 with a “hammerhead” shape that reached far above the other CDR H regions, explaining, in part, the potency of these two bNAbs [145]. The same group subsequently identified the PGT series of mAbs that shared similar glycan-binding specificities and the same high degree of potency and breadth as PG9 and PG16 [146]. More recently, bNAbs targeting the V1/V2 loop domain and epitopes that span both the gp120 and gp41 subunit have also been discovered [140, 147-151].

Collectively, these studies reveal that despite the inherent features of the Env trimer that serve to reduce antibody targeting, antibodies with potent neutralizing activity do evolve within infected hosts, albeit if in only 10-20% of HIV-1 infected patients [139]. The maturation of antibodies that target the recessed CD4 binding site or incorporate N-linked glycans into their epitopes likely reflects constant antibody stimulation through continuous viral replication of HIV. However, HIV-1 infected patients universally escape neutralizing antibody regardless of the potency of neutralization of the host. The most recent phase III human clinical trial, the RV144 “Thai” trial, found a 30% protection rate comparing the placebo group with those that received the prime/boost regime [152, 153]. However, they found no correlation of the partial protection with NAbS but instead found a correlation with nonneutralizing Env-binding antibodies that target the V2 loop [154, 155]. Similarly, live-attenuated strains of SIVmac239 that provide sterilizing protection in rhesus macaques, but analysis of antibody responses fails to detect potent NAbS in the protected animals [156-158]. Thus, the rare nature of bNAbs raises questions as to the role of these antibodies relative to the total Env-specific population of antibodies and whether efforts to induce responses that produce bNAbs are possible using current vaccine strategies.

The importance of antibody-mediated B-cell responses to Env in immunological control of SIV and SHIV

The SIV/macaque model of HIV/AIDS pathogenesis has been instrumental in our understanding of the antibody response to Env, through the use of sequence-defined pathogenic strains of SIV and the controlled setting of experimental infection (with defined doses and routes of inoculation, using immuno-
genetically defined Asian-origin macaque hosts) [159, 160]. SIV molecular strains, such as SIV\textsubscript{mac}239, and their derivatives provided a method to understand the importance of antibody responses in control of SIV infection in rhesus macaques. SIV\textsubscript{mac}239 is a molecular strain that is highly pathogenic in rhesus macaques and typically neutralization resistant to most SIV\textsuperscript{+} sera/plasma or mAbs specific for SIV Env [110, 161-163]. Site-directed mutations were engineered into SIV\textsubscript{mac}239 to address the effect of modifications within Env on neutralization sensitivity and replication during infection of rhesus macaques. SIV\textsubscript{mac}239 g45, g46, and g56 were mutated to remove N-linked glycosylation (NLG) motifs at the 4\textsuperscript{th}, 5\textsuperscript{th}, or 6\textsuperscript{th} N-linked glycosylation sites (NLGS) in and around the V1 loop of gp120 [164]. The NLGS were disrupted by mutating the conserved asparagine to glutamine in the highly conserved NXS/T motif. SIV\textsubscript{mac}239 M5 lacked 5 N-linked glycosylation sites (NLGS) in both V1 and V2. This panel of NLGS mutant derivatives were engineered based on the observation that N-linked glycan modifications played a critical role in immune evasion to antibodies by forming a “glycan shield” around the external surface of the Env trimer, but had yet to be verified in vivo [165].

Infection of rhesus macaques with these SIV\textsubscript{mac}239 mutant derivatives resulted in dramatically reduced viral load set points with SIV\textsubscript{mac}239 M5 replicating near or below the level of detection in complete contrast to SIV\textsubscript{mac}239 [163, 166]. Interestingly, NLGS mutant strains replicated similarly to SIV\textsubscript{mac}239 over the first two to three weeks of viral replication, even reaching peak viral loads near that of WT. Thus, there was no inherent defect in the replicative capacity of these strains due to the lack of any of the NLGS, suggesting attenuation was the result of increased sensitivity to Env-specific antibody.

Interestingly, macaques infected with SIV\textsubscript{mac}239 NLGS mutant forms produced more robust neutralizing antibody responses to both the mutant strains and to the parental SIV\textsubscript{mac}239 [166]. These studies further showed that the sera from animals infected with these mutant strains bound to V1 peptides spanning these NLGS to a much higher level than SIV\textsuperscript{+} sera from SIV\textsubscript{mac}239-infected animals, providing evidence that antibodies targeting epitopes shielded by these NLGS did emerge and potentially contributed to the increased neutralization titers against both the parental strains and the mutant NLGS strains. Furthermore, the use of strains that were isogenic except for the mutations removing NLGS in
these studies provided unambiguous evidence from an in vivo primate model that NLGS in gp120 functioned to shield Env from antibodies targeting epitopes near the vicinity of the attached carbohydrates.

A subsequent study again took advantage of the controlled setting of experimental SIV infection of rhesus macaques to further delineate antibody-mediated immunological control of viral replication. An additional artificially derived SIV\textsubscript{mac}239 strain with a 100 amino acid deletion encompassing the entire V1 and V2 loops (SIV\textsubscript{mac}239 ΔV1V2) in addition to SIV\textsubscript{mac}239 M5 were used to infect rhesus macaques [163]. Similar to SIV\textsubscript{mac}239 M5 the ΔV1V2 strain resulted in viral load set points below the level of detection, yet replication of both mutant strains was not distinguishable from SIV\textsubscript{mac}239 two weeks into infection during peak of viral loads. To define the level of immunological control of viral replication due to antibody another cohort of macaques were depleted of B cells by treatment with an anti-CD20 mAb over a period of 3 weeks. The animals were inoculated with SIV\textsubscript{mac}239 M5 during the 2\textsuperscript{nd} week of B-cell depletion. A subset of SIV\textsubscript{mac}239 M5-infected animals that were depleted of B cells showed a marked absence of viral control with viral loads at a level typical in SIV\textsubscript{mac}239-infected animals. Animals that received an isotype-control displayed the lower viral load set points observed in previous animals infected with SIV\textsubscript{mac}239 M5. Collectively these studies provided evidence that antibody responses impose a significant selective pressure on the replicating SIV population. Changes in Env that incur increases in sensitivity, either through the absence of variable loops or removal of NLGS, resulted in immunological control of viral replication not observed with the parental SIV\textsubscript{mac}239 strain [163]. Minimizing the selective pressure imposed by antibodies through B-cell depletion allowed the neutralization-sensitive strain SIV\textsubscript{mac}239 M5 to replicate to titers typical of parental SIV\textsubscript{mac}239 during the post-acute phase of viral replication [163]. In addition to confirming the importance of the antibody based B-cell immune response to primate lentiviruses, these in vivo studies highlighted the immune evasion properties of Env that play a significant role in the persistence of primate lentiviruses, the maintenance of high viral loads over the course of infection, and disease progression.

\textit{Env-mediated mechanisms of immune evasion from antibodies}
The following sections will review some of the inherent features of the Env trimer complex that function to reduce antibody recognition followed by the role of env sequence variation in antibody escape. 

*Two-receptor entry mechanism.* Both HIV and SIV use CD4 as a primary receptor and multi membrane-spanning chemokine molecules as secondary receptors during infection. As mentioned in the previous section, the initial interaction between gp120 and CD4 results in conformational rearrangements in gp120 that expose the coreceptor binding site. Coreceptor binding leads to dissociation of gp120 and exposure of gp41, effectively releasing the fusion peptide in a reaction that ultimately leads to fusion. The use of two receptors is not common among retroviruses [79]. Therefore, it is reasoned that PLV coreceptors were likely the primordial receptor and PLVs evolved the gain of interaction with an additional receptor (CD4) allowing Env to adopt conformations that occluded access to interactions that actually trigger fusion until absolutely necessary, effectively reducing exposure to antibodies [79]. Evidence supporting this model include the discovery of strains of HIV and SIV that could mediate entry by engaging coreceptor molecules without prior interaction with CD4 [109, 167, 168]. These CD4-independent strains typically gain tropism for macrophages in the central nervous system and brain, but are very sensitive to antibody-mediated neutralization, further supporting arguments that a two-receptor entry mechanism is an immune evasion strategy [110, 169]. A two-receptor mechanism functions to occlude the coreceptor binding site from antibody targeting, but essentially produced another target, the primary receptor binding site. The highly recessed cavity of the CD4 binding site may have evolved as a result of selection of antibodies that target the primary binding site, evidenced by the potency of neutralizing antibodies that can attain access to targets that overly the CD4 binding site.

Passage of HIV and SIV in tissue culture typically results in “lab-adapted” strains that are more neutralization sensitive when compared to the parental virus before passage. These lab-adapted strains can infect cells that express coreceptor molecules and little to no CD4 unlike the parental strains. Taken together this indicates that the necessity of CD4 to mediate entry is not absolute but does provide a higher level of neutralization resistance. The absence of antibody selection in cell culture results in viruses that
adopt conformations that increase exposure of the coreceptor binding site similar to CD4-independent strains that evolve to infect macrophages in tissue compartments with little to no antibody present.

**Occlusion of conserved regions of gp120 through higher-order oligomerization.** Env exists as a trimer complex on the surface of virus-producing infected cells and virions. An early prediction of the trimer complex was that oligomerization functioned to occlude access to the receptor binding sites in gp120 and conserved regions of gp41. Initial evidence supporting this hypothesis came from antibody mapping on the gp120 surface using both gp120 monomers and gp140 soluble trimers [170, 171]. These studies found antibodies that neutralized HIV-1 could bind to Env trimers, in contrast to non-neutralizing antibodies, which could bind gp120 monomers, but not trimer complexes. Antibody competition experiments found that neutralizing antibodies competed for a similar target on gp20, but did not compete against non-neutralizing antibodies. It was proposed that gp120 had a neutralizing face that was accessible on the trimer complex and a non-neutralizing face that was occluded through gp120 interactions in the trimer. Structural studies of gp120 monomers and gp140 trimers confirmed these early antibody-mapping studies. Mapping residues necessary for binding to neutralizing and non-neutralizing antibodies onto the first crystal structure of a HIV-1 gp120 monomer found that residues targeted by non-neutralizing antibodies mapped to a single domain, described as the inner domain based on the prediction that it lined the internal axis of the trimer, whereas the neutralizing face mapped to a region that spanned the inner and outer domains [87, 172]. The HIV-1 gp120 structure also confirmed that the CD4 binding site occurred within a deep cavity of gp120. A crystal structure of HIV-1 gp120 that contained the V3 loop suggested that the coreceptor binding site occurred within a pocket near the base of the V1/V2 loop and the V3 loop [90]. Cryo-electron microscopy studies of gp140 trimer complexes, and more recent crystal structures of highly engineered gp140 trimers, confirmed that the inner domain of gp120 was buried in the internal axis of the trimer and the non-neutralizing face of gp120 occurred along monomer:monomer interaction surfaces [114, 173].

**Glycosylation of the gp120 and ectodomain of gp41.** Glycosylation is a post-translational modification that attaches chains of carbohydrates to proteins. Env is modified by asparagine-linked (N-
linked) carbohydrates and O-linked carbohydrates which are attached to hydroxyl groups of serine or threonine residues [174]. N-linked glycosylation (NLG) occurs at asparagine residues within Asp-X-Ser/Thr (NXS/T) sequence motifs, where X can be any residue except for proline. Most HIV and SIV gp120 subunits have an average of 24 N-linked glycosylation sites (NLGS) and an average of 3 NLGS in the ectodomain of gp41. The sequence motif for O-linked carbohydrate attachment is not as well defined, but is predicted to occur in repeated stretches of serine and threonine residues. env sequences isolated from HIV-1 infected patients and SIV-infected rhesus macaques often contain mutations that introduce novel NLGS, remove NLGS, or shift NLGS (introduction of a mutation that knocks out an existing NLGS and introduces a novel NLGS). The first evidence that NLG was a mechanism of immune evasion came from a report that found a prevalence of mutations at NLGS in primary HIV-1 env sequences [165]. Testing Env-pseudotyped viruses with early and late timepoint env clones against matched longitudinal plasma identified the gain of neutralization resistance over the course of infection that was hypothesized to be due to the changes at sites of NLGS. The high density of NLGS in HIV-1 Env and, potentially, escape from antibody through mutations in NLGS led to the proposal of an evolvable glycan shield that functioned to protect Env from neutralizing antibodies [165]. They tested whether sequence variation at NGLS was responsible for viral escape by introducing NLGS changes observed late in infection into clonal isolates from early in infection. These NLGS mutants gained neutralization resistance against longitudinal plasma samples providing evidence that NLGS did shield Env from neutralizing antibodies and changes in NLGS were selected for within the viral population leading to viral escape. More recently, the identification of bNAbs that target the glycan modifications on the gp120 surface including 2G12, PG9, PG16, and the PGT series indicate the constant presence of the glycan shield on the gp120 surface led to the targeting of the glycans themselves [134, 140, 146].

Evidence for O-linked glycosylation also came from the SIV/rhesus macaque model. A study from the Overbaugh group used molecular clones from SIV of pig-tailed macaques, SIVmne, to show differences in neutralizing antibody sensitivities in parental SIVmne strains and variants with mutations at predicted sites of O-linked modification [175, 176]. This study found that plasma from SIVmne infected
animals could neutralize parental strains more potently than O-linked derivatives. Similar to the in vivo SIV\textsubscript{mac} studies, the only modification in virus sequence were those altering sites of O-linked glycosylation. Therefore changes in neutralization sensitivity could be attributed directly to the effect of O-linked modifications on antibody binding.

\textit{Variable loops}. Structural studies of gp120 find the V3 and V4 loops present on the external exposed surface of gp120 [87, 90, 177]. Although no structural information exists for the V1 or V2 loops within the context of the Env trimer, the base of the loops within monomer and trimer complex structures place V1 and V2 at the external surface of the gp120 monomer and near the apex of the Env trimer [173, 178]. Structural studies of gp140 trimer complexes and native cell-surface expressed Envs found conformational rearrangements of the V3 loop and the bases of V1 and V2 upon interaction with sCD4, indicating that the variable loops themselves function to block access to receptor binding sites on gp120 [50-52]. Further evidence for a role in hindering access comes from Env mutants that deleted one or more of the variable loop regions. For both HIV-1 and SIV removal of the V1/V2 loop renders virus severely neutralization sensitive, however, these variable loop deletion mutants are only partially replication deficient. SIV infection studies using a V1/V2 deletion mutant of SIV\textsubscript{mac}239 confirmed that removal of V1/V2 did not hinder replicative capacity but lead to complete control of viral replication by the host immune response in vivo and severe neutralization sensitivity to SIV\textsuperscript{+} plasma from SIV\textsubscript{mac}239 ΔV1/V2-infected animals when tested in neutralization assays [163].

1.5 \textbf{The Evolutionary Dynamics of PLVs}

The viral sequence diversity of primate lentiviruses, specifically in the variable regions of the \textit{env} gene, is a result of replicative features of PLVs coupled to constant immune selection. The dynamics of primate lentiviral infection under constant host selection is the reason HIV is one of the fastest evolving biological entities with some of the strongest positive selection observed in any organism [46]. Viral sequence diversity of PLVs is rooted in the rapid replication kinetics of the virus and the low fidelity rate, in addition to recombination properties, of RT. Viral escape by HIV and SIV is well established to both antibody and CTL adaptive immune responses, however, escape differs between the two responses.
because of differences in the nature of antibody recognition of viral proteins versus TCR recognition of viral peptides bound to MHC.

Replication kinetics of HIV and SIV. Two early landmark studies used similar approaches to estimate the rate of HIV-1 replication in vivo and the level of turnover of CD4+ T lymphocytes, the primary target of HIV and SIV [179, 180]. Using primary data from a pre-clinical drug study each group had access to viral load data before and during the initial treatment with an anti-viral drug. Using viral load data over the course of the treatment each group measured the rate of viral load decline to estimate the turnover rate for HIV-1 in vivo. Both studies estimated in vivo HIV-1 turnover rates near 1-2 days. The Wei et al study additionally estimated HIV-1 turnover using the rate of emergence of RT escape mutants to the protease inhibitor drug, which yielded similar rates of turnover. Using a similar modeling strategy with the rate of CD4+ lymphocyte increase during the course of treatment both studies yielded rates of CD4+ T cell turnover (production and destruction) around 2 x 10⁹ cells per day. These early studies set the precedent for the astonishingly high rate of HIV-1 virus replication and the degree to which they cleared their target cell population.

Error rate of PLV RT. The biological basis for much of the genetic variation observed in HIV and SIV is rooted in the rate at which RT incorporates an incorrect nucleotide during reverse transcription of the viral genome during one replication cycle or error rate and the rate at which RT switches templates during replication or recombination. RT is estimated to make ~0.2 errors per genome during each replication cycle [181]. RT has an error rate that falls between the highest error rates of RNA virus RNA-dependant RNA polymerases and the lowest error rates observed in DNA polymerases of DNA viruses, prokaryotes, and eukaryotes [182]. The higher error rates of RT and RNA virus polymerases is attributed to the lack of proofreading activity in these enzymes.

Recombination rate of PLVs. The estimated rate of HIV-1 recombination predicts three recombination events occur during viral replication [183]. This recombination rate is one of the highest recombination rates estimated for any organism [46]. Additionally, the recombination rates of HIV-2 and SIVsm, are estimated to be within the same range, thus the high rates of recombination is likely to be a
common feature of primate lentiviruses [184]. Several lines of evidence provide further support for the role of recombination in generating genetic diversity including the observations of dual infection (superinfection) [185, 186], the presence of more than one provirus in virally-infected cells [187], evidence of recombination within and among HIV-1 subtypes [188-190], and recombination generating novel viral lineages that possess mosaic genomes of their ancestors [34, 191]. Furthermore, recombination is the primary driving force behind the 15 different circulating recombinant forms of HIV-1 produced through inter-subtype recombination, which in some regions account for at least 25% of all HIV-1 infections [192]. Although not as significant, viral sequence variation is also predicted to occur from transcription of the proviral genome by host RNA polymerase II [193].

Mutation rate of PLVs. Based on the high error rate and high recombination rate of RT it was not surprising that HIV and SIV are estimated to have a high mutation rate. Initial mutation rates for HIV-1 were estimated using purified HIV-1 RT. Subsequent studies developed a cell-based system of HIV-1 infection limited to a single round of replication estimating the HIV-1 mutation rate to be \( \approx 3.4 \times 10^{-5} \) [194].

The high mutation rate of PLVs coupled to their rapid replication kinetics is proposed to lead to the generation of a population of genetically distinct viral variants within a given host. Eigen coined the term quasispecies to describe populations of distinct but related genomes that are subject to genetic drift and natural selection [195, 196]. Thus, evolutionary theories formulated under population genetics were fit to describe the evolution of RNA viruses, including primate lentivirus populations under the stochastic environment of transmission and the selective environment, including immune selection, of the infected host.

Intra-host and inter-host evolution in the env gene of HIV and SIV

Indeed the rate of intra-host \( env \) evolution is evidence by the level of immune selection mounted against \( env \) leading to positive selection [197-199], which in one study was estimated to produce at least one adaptive fixation event every 2.5 months [200]. The diversity found within SIV \( env \) is also though to account for the ability of SIVs to adapt to new host species [201, 202].
In contrast, the rate of inter-host evolution, mostly through sequence analysis of HIV-1, finds a different pattern of evolutionary dynamics [46]. Whereas comparison of env sequences isolated from individuals reveals positive selection, comparison of env sequences among different individuals shows little to no continuous positive selection of specific lineages (variants) [203]. Interestingly, it has been proposed that over the course of the HIV-1 pandemic host antibody responses have collectively selected for neutralization resistance and Env features, such as increased length of variable regions, that would function in antibody resistance [204-206].

**Genetic bottleneck during mucosal and intra-venous transmission of HIV and SIV**

The stark absence of positive selection from inter-host sequence analysis is mostly due to the genetic bottlenecks known to occur during PLV transmission. The application of single genome amplification (SGA) revealed that in many instances of primary HIV-1 transmission all variants found at the earliest stages of infection could be traced to a single infecting lineage that established the infection, referred to as the transmitter/founder (T/F) lineage or variant [207-209]. Studies using the SIV/macaque model supported these observations using the genetically defined SIV\textsubscript{mac}251 swarm [210]. Collectively, these studies helped to establish the paradigm of the bottleneck during HIV/SIV transmission resulting in a homogenous population during acute to early infection, the onset of host immune responses selecting upon the population leading to the emergence of sequence diversity, allowing for immune escape, changes in cellular tropism, and in the case of ARV-therapy, development of drug resistance.

**Evolutionary framework for estimation of HIV and SIV viral population sizes**

The observed intra-host diversity of HIV and SIV holds significant implications regarding the ability of HIV and SIV to generate escape variants to antiviral drugs and host immune responses. The high mutation rate and rapid replication of PLVs becomes intriguing when thought of in evolutionary terms. The theoretical framework for the evolution of populations is housed between random genetic drift and deterministic selection in the presence of selective forces. It is predicted that when populations are small (or selective forces are weak) that genetic drift dominates the fixation of mutations within the population, however, when population sizes are large mutations occur more frequently and natural
selection of mutants/variants within the population begins to dominate leading to a deterministic regime of evolution in the population (reviewed in [211]). One would initially presume that natural selection of most fit variants would likely dominate the evolution of genetically diverse HIV and SIV populations with such a high mutational rate, however additional factors that effect population sizes do not completely rule out scenarios where random drift dominates. For instance, some mutations yield non-viable progeny virions that would not contribute to the effective population or $N_e$. Additionally, variation in replication and generation time would yield smaller population sizes then would be predicted from the number of infected cells.

Therefore, the evolution of HIV and SIV likely occurs through a dynamic regime of drift during acute infection but becomes deterministic as the population grows. The critical value at which random drift begins to give way to natural selection is predicted to occur when the population size reaches a number that is the inverse of the mutation rate of the organism [182]. This prediction led to several attempts to define the population sizes of HIV within infected hosts to better understand the evolutionary dynamics of the virus under stochastic or deterministic models. Initially several studies estimated relatively small population sizes [212-216], suggesting stochastic models would best fit viral dynamics of HIV-1, however, more recent analysis have estimated effective population sizes near $1.5 \times 10^5$, in an average infected patient [217-219]. This range is far below what would be expected based of the mutation rate and replication kinetics of the virus, but it is within a range where mutations are maintained long enough for selection to occur [217]. Thus deterministic models of evolution are likely to be more accurate for HIV and SIV.

**Escape from adaptive cellular-mediated immune responses.**

CD8$^+$ cytotoxic T lymphocytes (CTLs) are constantly surveying for the presence of pathogens through their antigen specific T-cell receptor (TCR), which interacts directly with cell surface expressed MHC class I molecules loaded with peptides. Immune evasion from MHC class I-restricted CTL responses occurs through Nef-mediated down regulation of MHC class I molecules in addition to sequence variation leading to escape [220]. During synthesis and trafficking to the cell surface MHC
class I molecules are loaded with peptides that are derived from host cell proteins (self peptides) or, in the case of a virally-infected cell, are derived from proteins of viral pathogens (viral peptides). CTL recognition of a MHC class I molecule loaded with a viral peptide on the surface of an infected cell leads to CTL-mediated lysis of the infected cell through the release of perforin and granzyme [221, 222].

Viral peptides may be generated from any degraded viral protein, therefore CTL recognized epitopes do occur in functionally or structurally conserved regions incurring a steep fitness cost to viral escape. The viral peptides loaded onto MHC class I molecules are ~8-9 amino acids in length. MHC class I molecules bind near the ends of the peptides, thus viral escape from CTL-mediated responses typically occurs through changes that reduce binding and loading of peptides unto MHC class I molecules and are typically observed at the 2\textsuperscript{nd}, 8\textsuperscript{th}, and 9\textsuperscript{th} positions of the linear viral peptide. Defining the peptide binding specificities of MHC class I molecules by \textit{in vitro} peptide binding assays helped to identify putative CTL epitopes and identify sites of viral escape.

MHC class I molecules are not antigen specific like TCRs and antibodies. Humans can express up to six alleles of MHC class I molecules, and even more in primates, increasing the breadth of peptides bound for presentation to TCRs in a given host [223]. In PLV infection, the MHC haplotype dictates the epitopes targeted by CTL and escape changes selected for in the host [224, 225]. MHC molecules in humans and rhesus macaques that restrict peptides from functionally or structurally conserved regions are associated with delayed disease progression [226-230] and positive disease outcome [231, 232].

CTL epitopes in highly conserved or functionally constrained regions elevates the fitness cost of escape, making CTL immune targeting a potent selective force [233, 234]. CTL responses during acute infection are one of the primary selective forces driving viral loads down from their peak to set viral loads [5, 235]. However, once escape mutations emerge, MHC recognition and CTL targeting of the epitope is lost. In HIV and SIV infection escape from CTL may be generated rapidly or late in the chronic phase of infection leading to a loss of viral control [234, 236-239], or in some instances escape may lead to subsequent immune control [225, 240]. Additionally, compensatory mutations that decrease the fitness...
burden incurred by the escape change function to maintain replicative fitness under CTL immune surveillance [241, 242].

In summary CTL adaptive immune forces are a potent, effective mechanism of viral control. However, the generation of viral diversity leads to rapid viral escape leading to a loss of viral control. A detailed understanding of MHC class I restriction and the linear nature of CTL epitopes helped identify sites of CTL escape. Subsequent studies defined the fitness costs they incurred and identified compensatory changes that relieved the fitness burden of escape changes, increasing the overall fitness of the viral population to maintain persistence.

*Escape from antibody-mediated humoral responses*

Antibody responses to PLVs are primarily mounted at the surface exposed envelope glycoprotein. Antibodies, in contrast to CTL recognition of MHC restricted linear peptides, recognize both linear and conformational epitopes. Sites within a single antibody conformational epitope may exist in regions of Env that are distal to each other in the polypeptide chain, but are brought together in the secondary, tertiary, or even quaternary structure of the gp120 subunit, making mapping of antibody escape difficult to ascertain. Additionally, mapping of antibody targets with monoclonal antibodies found that changes in one region or subunit of Env could affect binding to a distal region or subunit of Env. For instance, changes in the gp41 subunit resulted in reduced binding to mAbs that recognized the gp120 subunit [243-245]. Therefore, changes outside of the actual epitope could lead to escape, likely by introducing conformational changes that reduced access to the target epitope.

Initial reports of SIV *env* sequences found the presence of conserved regions interspersed with variable regions similar to HIV. An advantage of the SIV/macaque model was the ability to compare *env* sequences that evolved in SIV-infected macaques to the parental strain used to infect the animals providing direct sequence evidence of sites of adaptation. SIV *env* variation was reported from experimental infection of rhesus macaques with SIVmac239 and experimental infection of pig-tailed macaques with SIVmneCL8 [246, 247]. Both studies found a significant amount of genetic variation located in variable regions similar to HIV-1 *env*. Burns *et al.*, measured variation from 27 primary *env*
clones isolated from two SIV\textsubscript{mac}239-infected rhesus macaques. \textit{env} clones were isolated from chronic timepoints (week 69 and 93) in one macaque and from week 43 p.i. in the second macaque. Comparing all of the nucleotide substitutions relative to the total number of sequenced nucleotides they found over 81\% of all changes were non-synonymous changes, and strikingly over 98\% of changes in variable regions resulted in an amino acid change. Based on the time of infection when the \textit{env} sequences were isolated the estimated rate of nucleotide substitution was $8.5 \times 10^{-3}$ per site per year. This rate was similar to early estimates of the HIV-1 nucleotide substitution rates and the rates for SIV and HIV-1 were both more then a million times higher than rates estimated for mammalian genomes [246, 248]. Very similar observations were found in the \textit{env} gene with the SIV\textsubscript{mac}/pig-tailed rhesus macaque primate model. Again, over 98\% of nucleotide changes in variable regions resulted in an amino acid change. Interestingly, both models found very little variation in the V3 loop of SIV \textit{env}, relative to the corresponding V3 loop of HIV-1.

Even though each study used different molecular clones of SIV and different primate hosts, substitutions were shared between the two \textit{env} genes. In-frame deletion and insertion (indels) events were observed in the V1 loops of both SIVs. Many of the indels in V1 were within a Ser/Thr rich stretch of residues. Substitutions in V1 were also concentrated within the Ser/Thr repeat region and typically involved Thr $\rightarrow$ Ser/Ala/Ile/Pro residue changes, Ile $\rightarrow$ Thr/Ala changes, and Pro $\rightarrow$ Ser changes. Among all these substitutions, typically only a single nucleotide change is required for the amino acid change.

In-frame deletions were observed in 11 of the 27 \textit{env} sequences in the SIV\textsubscript{mac}239 V4 loop spanning amino acid positions 416-425 [246]. Substitutions observed in the V4 loops of both SIVs included substitutions introducing a NLGS at the same site, D415, and substitutions at P421 including Pro $\rightarrow$ Leu/Thr/Gln. The observation of similar changes within corresponding regions of SIV during \textit{in vivo} replication of two similar but genetically out-bred primate hosts is strong evidence of a pattern of viral variation that occurs along similar genetic pathways in response to a strong selective force. Of note, these adaptations are consistently reported in SIV\textsubscript{mac}-infected Asian-origin macaque species [249-257]. SIV\textsubscript{mac}
infection of Asian-origin rhesus macaques represents a recent cross-species transmission event in which the virus has adapted to the new host in a relatively short period of time (~30 years), however, reports of env sequence (focusing on the V1/V2 region) in the context of natural SIV<sub>sm</sub>-infection of sooty mangabey monkeys revealed the same pattern of substitutions in the Thr/Ser rich region of V1 and similar levels of genetic variation and positive selection [258]. SIV infection of African primates represents millions of years of co-evolution. Therefore, the genetic patterns observed in the SIV/macaque animal model reflect the ancient SIV lineage and its long history of replication and adaptation in the primate host.

A subsequent study of an SIV<sub>mac</sub>239-infected rhesus macaque was able to directly connect Env adaptations to escape from antibody based neutralization targeting SIV<sub>mac</sub>239 Env [162]. Similar to reports of HIV-1 infected individuals with rare potent neutralizing activity, this SIV<sub>mac</sub>239-infected rhesus macaque possessed potent neutralizing activity of the parental SIV<sub>mac</sub>239 at a 1:1000 fold dilution, a titer never before reported in the literature. The significance of this finding still stands, as an antibody has yet to be isolated from SIV-infected macaques or HIV-infected humans, that can potently neutralize SIV<sub>mac</sub>239, with only exception being a bi-mimetic peptide engineered to bind the CD4 binding site and the CCR5 co-receptor binding site [259].

Testing longitudinal samples of plasma from Mm 333 found potent neutralization at 16 weeks post infection (p.i.) that developed some time between 4 and 16 weeks p.i. Interestingly, the viral load course for animal Mm 333 was typical of SIV<sub>mac</sub>239-infected macaques, indicating viral escape had occurred from the potent neutralizing activity mounted by Mm 333. Primary env sequences spanning the gp120 and ectodomain of gp41 from 16 weeks p.i. and 42 weeks p.i. were isolated and cloned back into parental SIV<sub>mac</sub>239. Testing these primary Env clones against plasma from weeks 16, 42, and 88 of infection found increased resistance from all Env clones against week 16 plasma relative to parental SIV<sub>mac</sub>239. Substitutions found in the primary Env clones were then introduced individually into SIV<sub>mac</sub>239 and tested against week 82 plasma to identify individual sites responsible for increased neutralization resistance. These neutralization experiments found three substitutions in gp120 provided a ~2-log increase in neutralization resistance relative SIV<sub>mac</sub>239 and two substitutions in gp41 that provided
a 30-60 fold increase in neutralization resistance relative to SIV\textsubscript{mac}239. Of the three gp120 substitutions Env A138T existed in the Thr/Ser rich hypervariable region of V1 flanking Thr residues predicted to be attachment sites of O-linked glycosylation [260]. The other two substitutions mapped to the V4 loop. One substitution was at Env A417T, which introduces a NLGS at position 415. Of note, SIV\textsubscript{mac}239 does not possess any other NLGS in the V4 loop thus this substitution represents the emergence of a novel NLGS that likely contributes to the glycan shield encompassing the V4 loop. The third substitution was Env P421Q, a site with one of the highest levels of polymorphism observed in SIV\textsubscript{mac} env. Although, the two sites in V4 were within the range of a single antibody epitope Env A138T was in the V1 loop yet provided the same level of neutralization resistance as A417T and P421Q. It was proposed that the distribution of neutralization resistance over two distantly positioned variable loops was potentially occurring through global conformational changes as a result of the individual substitutions reducing access to the antibody or antibodies responsible for the potent neutralization of the polyclonal plasma of Mm 333. This phenomena was observed in HIV-1 Env when a single amino acid change at position 582 in gp41 resulted in increased neutralization resistance to mAbs targeting gp120 [243, 245, 261]. It still remains to determined what the target of the potent neutralizing activity of Mm 333 was and the mechanism of resistance induced by the individual changes in gp120.

This case study highlighted, once again, that despite the ability of the host to mount a robust and effective antibody response to the virus, Env adaptations led to viral escape that allowed for persistent infection and progression to disease. This study also found adaptive changes important for antibody escape mapped to variable regions, specifically V1 and V4, and to the ectodomain of gp41. Interestingly, this pattern of variation is observed in monkeys that possess high or moderate neutralization titers to the parental challenge strain or swarm providing evidence that the nature of viral escape occurs through commonly used mutational pathways to evade host antibodies regardless of the nature of the antibody response. Therefore understanding the nature of virus evolution during the acute and early stages when adaptive changes in response to antibody begin to emerge would provide insight into the nature of viral escape within the SIV/macaque primate model of AIDS pathogenesis.
Selection and fitness of viral adaptations that confer immune escape

Here, viral replicative fitness is defined as the ability of a variant to pass on its genetic material to progeny virions. Viral replicative fitness encompasses all stages of the virus cycle including entry, reverse transcription, integration, viral gene transcription and expression, assembly, budding of immature particles, and finally maturation [262]. Experimentally, the most straightforward measure of viral replicative fitness is the relative fitness of a virus mutant- typically measured against a genetically defined reference strain in a mixed infection, or competition assay.

Analysis of viral replicative fitness of HIV and SIV has focused on mutations that arise in response to antiviral drugs (reviewed extensively in [262-265]). Analysis of HIV-1 RT mutants that confer drug resistance were found to reduce viral replicative capacity when measured against molecular clone HIV-1 NL4-3 in dual (WT + mutant) competition infections [266]. Relative fitness was determined by quantitative real-time PCR (qRT-PCR) targeting an amplicon spanning an 11-nt silent modification introduced into \textit{vif}. Interestingly, mutants with an increased or decreased effect on fidelity had lower relative fitness to WT, suggesting the mutation rate of HIV-1 exists near an optimal threshold. Two recent studies have focused on the fitness cost to CTL escape in HIV-1. Boutwell \textit{et al} tested 3 CTL escape variants in a dual competition assay that utilized allele-specific qRT-PCR targeting a 4-nt silent modification sequence tag introduced into \textit{nef} and 29 CTL escape mutations in \textit{gag} by qRT-PCR using primers targeting a region within \textit{gag} that encompassed the mutations of interest [267, 268]. These studies found that individual mutations with \textit{gag}-targeted CTL epitopes did carry a replicative fitness cost, and certain combinations of mutations increased the total fitness costs. A viral fitness cost to CTL escape would be expected as CTL epitopes exist within regions, such as Gag, which must maintain sequence conservation to maintain structure critical to higher order multimerization. Although, escape from antibody would be predicted not to incur replicative fitness costs, especially antibodies that target the unstructured loop regions, few studies have addressed antibody fitness costs through quantitative measurements from competition assays. A recent study found escape from bNAb VRC01, which targets the conserved CD4 binding site, was associated with viral fitness loss, which would be expected as sites
of escape occurred in regions critical to CD4 binding [269]. However, even this recent study relied on growth curves from single virus infections measured by p24 antigen-capture ELISA. As we were interested in adaptations in variable loop regions, we predicted that mutations responsible for amino acid adaptations in variable loop regions potentially have little to no associated fitness cost. Therefore we wanted to address the relative fitness of mutations within SIV Env where adaptations repeatedly emerge using experimental conditions with the highest resolution to detect differences in replicative capacity.

Several methods to measure the replicative fitness of PLVs have been developed, but the best current method involves measuring the viral replication of the variant and the reference strain in a single mixed infection or competition [263]. A competition of the mutant and the reference strains in a single well provides an internal control for any variability that may arise when measuring replication of the two variants separately. Despite the advantages of mixed infections, competition experiments introduce another experimental consideration, which is the method used to distinguish the WT and mutant strains, which are essentially isogenic except for the mutation of interest. Several approaches exist to distinguish WT and mutant strains including peak heights in chromatographs [270], heteroduplex tracking assays [271], flow cytometry-based methods using GFP/RFP-labeled viruses [272], modified reporter gene systems [273, 274], or allele-specific quantitative real-time PCR [268, 275, 276].

Although, each of these methods provide quantitative in vitro measurements of relative viral fitness, they have two disadvantages: First, both of these methods require the introduction of sequence that alters the primary sequence, and these modifications are required for every variant to be tested; Second, most viral competition assays are limited to dual infections of a wild type (WT) reference strain and a mutant strain, thus limiting the number of mutations that can be screened in a single experiment. Here, we developed a method that addressed these two disadvantages, and, additionally, maintained a high resolution for detection of relative fitness differences in competition-based experiments. This deep sequencing based-viral fitness assay provides a method to test more then one mutant in a single experiment using viruses expressed from full-length plasmids without any modifications.
1.6 Scope of dissertation

It is well understood that PLVs produce populations of genetically distinct variants due to their high mutation and virion production rates and rapid replication, conferring the ability to evade robust, virus-specific adaptive immune responses. The adaptability of PLVs contributes to the rapid loss of immune-mediated viral control, the establishment of persistent infections, and the failure of vaccine-induced immune responses to provide protection from HIV-1. Therefore, characterization of viral evolution at the population level is critical to a fundamental understanding of the adaptability of primate lentiviruses to host immune responses.

The recent advances in sequencing technologies have brought in a new era in genomic sequencing. Next-generation sequencing (NGS) technologies are typically based on a platform that allows for parallel sequencing of multiple templates at one time producing millions of bases of sequence data per reaction. The advantage in NGS technologies over benchmark Sangar sequencing lies in the level of sequence information generated from a single sample. The depth of resolution of NGS can reach up to 10,000 fold coverage of the HIV-1 genome, or even higher when focused to a specific region or gene. This level of coverage provides a method of detection of minor or rare variants within the sample as would be predicted during the early stages of HIV or SIV infection providing critical information into the first adaptive events of the virus population that previously would be undetectable by standard sequencing.

The advantage of the SIV/macaque model is the use of well characterized pathogenic clones of SIV with defined sequences, such as SIV\textsuperscript{mac239}, or uncloned swarms, such as SIV\textsuperscript{mac251}, which provide the ability to define the consensus sequence and the diversity of the inoculum prior to experimental infection. Investigation of any longitudinal samples from SIV-infected macaques can then be tracked from the defined stock to unambiguously define adaptive changes during the acute and early stages of infection.
Here we apply deep sequencing approached to viral populations at the acute and early stage of infection in order to identify and better understand the initial adaptations selected for by antibody within viral populations.

This dissertation extends our understanding of the dynamics of SIV evolution focusing on the SIV envelope glycoprotein over 29 weeks of in vivo replication in the rhesus macaque host and extends our knowledge of the effects of antibody escape changes on viral replicative fitness. Understanding the viral fitness cost of adaptations critical for antibody escape could also guide which adaptations, if any, provide the largest barrier to escape in the host and would be important changes to test in Env-based immunogens using the SIV/macaque model of infection. In the third chapter of this dissertation, I describe several adaptations that emerged during early SIV infection in the V1 and V4 loops of the gp120 subunit and their effects on antibody binding, antibody neutralization, and antibody-dependent cellular-mediated cytotoxicity. I also used a novel fitness assay to measure the replicative fitness costs associated with a subset of these adaptive changes in Env. We found that Env substitutions that confer escape to high-titer neutralizing plasma do not incur a viral fitness cost, supporting a model in which the variable loops have been selected as a mechanism to carry the burden of constant antibody targeting while protecting conserved structural features responsible for mediating cellular receptor binding and fusion. Although this model would suggest an unlimited selection of substitutions, we find that variation is highly focused at specific sites within two clusters, spanning 8-10 residues in the V1 and V4 loop regions. The repeated emergence of the same adaptations in related SIV strains adapted to outbred macaque species and SIVs that naturally infect primate hosts suggests a pattern of variation that maximizes antibody escape without incurring a viral replicative fitness cost.

We extended the practicality of our deep-sequencing based viral fitness assay by asking whether a previously determined nonneutralizing Env-binding antibody could inhibit the replication of the typically neutralization resistant SIVmac239. Although these results are preliminary, they suggest standard neutralization assays using standard reporter cell lines do not have the sensitivity to detect subtle, but significant neutralization activity and are not recapitulating what may be occurring in vivo.
CHAPTER 2

Distinctive Patterns of Evolution Consistently Associated with

SIV_{mac}251 Env during Early Infection
2.1 Acknowledgments

Sergio Ita conceived of experiments, produced and prepared samples for Illumina sequencing, performed neutralization, ELISA-binding, and ADCC assays, collected and analyzed results, and drafted this text.

Dr. I.B. Fofana provided antibody binding and neutralization data in Figure 6.

Deep sequencing was performed at the Dana Farber Cancer Institute Genomics core. Variant calling and analysis was performed with the assistance of Ruchi Newman, Michael Zody, and Xiao Yang of the Broad Research Institute’s Viral Genomics Group.

Viral load data were generated by Jeff Lifson and Michael Piatek at NCI Frederick.

All animal work, sampling, and care was provided by the veterinary staff of the New England Primate Research Center.

This work was supported by grants AI083118 (NIH) to Dr. Welkin E. Johnson.
2.2 Abstract

Primate lentiviruses (PLVs), including human immunodeficiency virus type 1 and 2 (HIV-1 and HIV-2), and the simian immunodeficiency viruses (SIVs) cause persistent lifelong infections despite the presence of virus-specific adaptive immune responses. The target of antibodies is the viral envelope glycoprotein (Env), which is expressed on the surface of virions and infected cells as a trimer of gp120:gp41 heterodimers. While many studies have described HIV and SIV env sequence variation arising from evasion of antibodies, the dynamics of viral env gene evolution during the earliest stages of infection (< 30 weeks) have not been well defined.

To examine the evolution of Env during the earliest stages of infection, we obtained longitudinal samples spanning acute and early infection (2-29 weeks post-infection) from a cohort of four SIV-infected rhesus macaques, as well as the inoculum (a swarm stock of uncloned SIVmac251) used to infect the cohort. Despite the small number of animals, the cohort displayed a range of viral loads and disease outcomes: one animal maintained high viral loads throughout the course of infection and rapidly progressed to disease, one animal showed signs of viral control in the post-acute phase of infection, and the remaining two animals displayed typical viral load outcomes. Three of the four animals mounted detectable antibody responses to Env, measured from plasma binding to soluble SIVmac239 Env (gp140) by ELISA, while the animal with highest viral loads and rapid disease progression did not mount a detectable Env-specific antibody response. We used deep sequencing to analyze the inoculum and track evolution of the viral population, capturing a population bottleneck at the point of transmission from the stock into each animal and the subsequent emergence of Env diversity from the initially homogeneous population that correlated with the onset of Env-specific antibodies. We tracked the appearance and fixation of substitutions, insertions, and deletions in each animal and found that adaptive changes were tightly clustered in the V1 and V4 loops of the gp120 domain of Env. These patterns were strikingly similar between animals, and very similar to previously reported patterns for SIVmac, SIVmne, and SIVsm. Surprisingly, multiple large in-frame deletions in V4 emerged and dominated the viral population in two of the animals with detectable antibody responses to Env.
Adaptations at P421 in the V4 loop led to a loss of binding to several V4 loop-specific mAbs. P421Q increased the neutralization resistance of otherwise neutralization-sensitive SIV\textsubscript{mac}316 to all V4 loop-specific Rh mAbs. When we tested our cohort of animals for antibody-dependant cellular-mediated cytotoxicity (ADCC) we found that animals with detectable Env-binding antibodies had ADCC activity against SIV\textsubscript{mac}239 when measured using a terminal-timepoint plasma. Mm 156-10, which did not have detectable Env-binding antibodies similarly had no ADCC activity against SIV\textsubscript{mac}239. Interestingly, neither neutralization escape changes nor in-frame multiple amino acid deletions conferred resistance to ADCC activity in our animals, suggesting escape to ADCC may require multiple changes within Env.

Altogether, the deterministic appearance of changes leading to replacement of the residue present at transmission, polymorphism, and insertion/deletions revealed a remarkable degree of viral protein evolvability. Moreover the observation that similar or identical substitutions, insertions, and deletions occur in multiple primate species infected with related but divergent strains of SIV revealed a reproducible pattern of Env variation, suggesting constant antibody targeting results in predictable, specific pathways of viral escape during the early stages of infection that mediate loss of antibody recognition and neutralization resistance.
2.3 Introduction

The primate lentiviruses (PLVs), including the type 1 and type 2 human immunodeficiency viruses (HIV-1 and HIV-2) and the simian immunodeficiency viruses (SIVs) cause persistent infections involving continuous, ongoing viral replication that can last from months to years in the infected host. The ability of PLVs to replicate continuously in the face of innate and virus-specific adaptive immune responses stems from a complement of different immune evasion strategies [79, 80].

The envelope glycoprotein (Env) of HIV and SIV, which is expressed as a trimeric complex on the surface of virions and infected, virus-producing cells, serves as the primary target of antibody-based immune responses. Env mediates entry through direct interactions with host cellular receptors followed by fusion of the viral membrane with the host cell membrane. The primary receptor for both HIV and SIV is the CD4 cell surface molecule and the predominant co-receptor molecules are the multi-membrane spanning chemokine molecules CCR5 and CXCR4. During synthesis the Env precursor protein (gp160) trimerizes, is modified with glycans within the ER and is then cleaved in the Golgi to produce the surface subunit (gp120) and the transmembrane subunit (gp41), which remain non-covalently attached. Together the gp120:gp41 heterodimer forms a single unit of the trimer complex.

Rhesus macaques infected with pathogenic SIV strains develop continuous and robust antibody responses to Env, in a manner similar to HIV-1 and HIV-2 infected patients [119, 162, 277, 278]. While, Env-binding antibodies are mounted against both the gp120 and gp41 subunits, most SIV-infected macaques only mount low to undetectable neutralizing antibody responses to the parental strain over the course of infection. Similarly, in HIV-infected patients neutralizing antibodies are generally low in potency and narrowly strain-specific in breadth early in infection [119-121, 165, 279, 280].

Studies of HIV-1-infected patients with high titer neutralizing antibody responses led to the isolation of neutralizing antibodies that possess high potency and breadth when tested against large panels of viruses pseudotyped with primary HIV-1 Env clones [140, 146, 281-283]. These potent, broadly neutralizing antibodies target highly conserved targets on Env including the CD4-binding site [133, 141, 142], glycan-dependant epitopes in variable loop regions including the V1/V2 loops and the base of the
V3 loop [134, 140], quaternary epitopes present only on the fully formed trimer complex [146], and epitopes that span both the gp120 and gp41 ectodomain surface [149, 151]. More recent studies have identified antibodies in virally-infected hosts with alternative effector functions, most notably antibody-dependent cellular mediated cytotoxicity, that do target PLV Env [284-287], potentially driving sequence variation in Env as well.

PLV have evolved several mechanisms to evade Env-specific antibody responses including occlusion of conserved receptor binding regions through oligomerization, extensive glycosylation, a two receptor (primary and co-receptor) mechanism of entry, and the presence of surface exposed loops that mask functionally conserved regions [79, 278]. All together these mechanisms appear to contribute to the low antigenicity of Env and the neutralization resistance observed in primary strains of HIV-1 and SIVmac, including the commonly used SIVmac239 and SIVmac251 strains. Despite the high level of neutralization resistance of these SIVmac strains, a significant level of sequence variation is observed in the env gene of SIVmac, again similar to the env gene of circulating HIV-1 [246, 247].

Antigenic escape from antibody responses is well documented, however understanding viral escape through sequence plasticity is confounded by the fact that antibody responses are polyclonal, recognition in some cases involve binding to discontinuous or conformational epitopes, and B-cell receptor affinity evolves by somatic hypermutation and continuous selection of the B-cell population.

In some cases amino acid changes that alter neutralization sensitivity do not occur within the recognized epitope [175, 243, 245, 261]. Therefore escape changes can function indirectly, potentially by changing Env conformations, leading to reduced accessibility to regions targeted by neutralizing antibodies, complicating our understanding of the nature of antibody escape.

Altogether, the effect of immune selection on PLV env is strikingly apparent. Both HIV env and SIV env sequence analysis reveal a high rate of positive selection, in which the per site rate of nonsynonymous ($d_N$) exceeds that of synonymous substitutions ($d_S$) [203, 246]. The extreme sequence plasticity of PLV Env is located in five variable regions (named V1-V5) that are interspersed between constant regions (named C1-C5) of Env. The V1-V4 regions are flanked by conserved cysteines that
form disulfide bonds, resulting in variable loop regions. The variable loops are present on the outward surface of gp120, potentially masking conserved inner and outer domains that contain receptor binding sites. The rate of positive selection is highest when focusing on changes in variable loop regions [246, 247].

A remarkable degree of positive selection, specifically in the V1 and V4 loops, was initially discovered by bulk PCR and cloning of env sequences isolated from chronically infected macaques [162, 246, 247]. Variation in V1 and V4 is similarly observed in other SIV/macaque models of infection, including SIV	mne infection of pig-tailed macaques and SIV	sm infection of rhesus macaques [249-256], as well as in the context of natural infection in sooty mangabeys [258]. However, the limited sampling of bulk PCR and cloning does not provide an accurate picture of env evolution at the viral population level, and very few studies have focused on early SIV infection.

Next generation sequencing (NGS) or deep sequencing of HIV and SIV have provided a method to study viral sequence dynamics to understand escape from cytotoxic T cell (CTL) responses to HIV-1 and SIV [288-291]. In this study, we sought to take advantage of the SIV/rhesus macaque model and the resolution of deep sequencing of longitudinal plasma samples from SIV-infected rhesus macaques to examine the evolution of SIV env during acute and early infection and track the emergence of adaptations in regions targeted by antibody. Specifically, we tracked the in vivo sequence evolution of the SIV env gene in rhesus macaques infected with an uncloned SIV	mac251 swarm stock using high-resolution Illumina deep sequencing of the SIV	mac251 stock inoculum and longitudinal samples ranging from 2-29 weeks post-infection from four SIV-infected rhesus macaques. We captured a population bottleneck at the point of transmission from the inoculum into each animal and the subsequent emergence of Env diversity from the initially homogeneous population. Furthermore, we tracked appearance and spread of specific substitutions and in-frame insertions/deletions in the viral populations in each animal and found that changes were highly concentrated in the surface-exposed V1 and V4 loops. We characterized a subset of these adaptations for their effects on antibody binding using a panel of Env-specific rhesus monoclonal antibodies (Rh mAbs). We found several adaptations at position 421 resulting in the removal
of proline led to a loss of antibody binding to V4-loop specific Rh mAbs and that one adaptation altered neutralization sensitivity to the same Rh mAbs in the context of fully replicating virus. Further, animals that mounted Env-specific antibodies had detectable ADCC responses when measured against SIV\textsubscript{mac}239. Interestingly, several adaptations including neutralization escape changes and in-frame multiple amino acid deletions in V4 did not confer resistance to ADCC activity of these animals.

In summary, evolution of changes leading to replacement of residues present at transmission, including both substitutions and insertion/deletions, illustrated a remarkable degree of viral protein evolvability, which manifested in a time-frame that correlated with the appearance of host Env-specific antibody responses. We found V4 loop adaptations that emerged to high frequency in the SIV populations of animals in our cohort and have been reported by others alter antibody recognition and neutralization. Further, these changes followed a selective pattern of sequence evolution verifying that the early antibody response poses a selective barrier on the viral population during early SIV infection
2.4 Materials and Methods

Viruses and plasma

All samples described in this study were obtained from archived material at New England Regional Primate Research Center. Rhesus macaque monkeys (Mm) 10-10, 198-08, 156-10, and 174-08 were unvaccinated control animals challenged with the uncloned SIV\textsubscript{mac}251 (swarm) as part of an unrelated vaccine study. The animals were housed at the New England Primate Research Center (NEPRC), and given care in accordance with standards of the Association for Assessment and Accreditation of Laboratory Animal Care and the Harvard Medical School Animal Care and Use Committee. All four animals were challenged with 6 doses of the SIV\textsubscript{mac}251 swarm at 300 50% tissue culture infective doses (TCID\textsubscript{50}). The SIV\textsubscript{mac}251 stock was obtained through the NIH AIDS Reagent Resource Program, Division of AIDS, NIAID, NIH.

Cell lines

The human embryonic kidney cell line 293T/17 (HEK 293T/17) was used for transient transfection of full-length viral plasmids for production of virus stocks of SIV\textsubscript{mac}239 and protein expression plasmids for recombinant SIV\textsubscript{mac}239 gp140 protein. HEK293T/17 cells were obtained from American Type Culture Collection (Manassas, VA) and maintained in D10 cell culture medium consisting of Dulbecco’s modified Eagles medium (DMEM) supplemented with fetal bovine serum (FBS) at 10% of the total volume. TZM-bl cells were derived from human HeLa cervical carcinoma cells and used for titration of viral stocks. TZM-bl cells were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: TZM-bl from Dr. John C. Kappes, Dr. Xiaoyun Wu and Tranzyme Inc. [292-295] and maintained in D10. Human CD4 \textsuperscript{+} C8166-secreted alkaline phosphotase (C8166-SEAP) cells are C8166 cells transduced with a plasmid encoding the alkaline phosphatase gene under the transcriptional control of SIV LTR sequences, as previously described [296]. C8166-SEAP cells were used for titration of virus infectivity, virus neutralization assays, and viral fitness assays. The cells were maintained in R10 consisting of RPMI medium supplemented with FBS at 10% of the total volume. Rhesus immortalized hybridoma B-lymphocyte cell lines (BLCLs) were derived by Rhesus Epstein-Bar
virus (EBV) transformation of peripheral blood B cells from rhesus monkeys infected with SIV_{mac} strains, as previously described [110, 297, 298]. These cells were the source of rhesus monoclonal antibodies that recognize SIV Env and were maintained in R20 supplemented with primocin. CEM.NKR-CCR5 CD4^{+} T cells are the target cell line for the antibody-dependant cellular-mediated cytotoxicity (ADCC) assay and were transduced with a pLNSX-derived retroviral vector to express Firefly luciferase under the transcriptional regulation of the SIV LTR promoter, as previously described [299]. The target cells were maintained in R10 supplemented with primocin. The effector cells for the ADCC assay are the human NK KHYG-1 cell line which express rhesus CD16 and were maintained in R10 supplemented with 1 μg/ml of cyclosporine (CsA) and 10 U/ml of interleukin-2 (IL-2), as described previously [299]. Both the target and effector cell lines were a kind gift from Dr. David Evans.

**Viral RNA isolation**

SIV viral RNA (vRNA) was extracted from plasma of SIV_{mac}251-infected rhesus macaques and the uncloned SIV_{mac}251 stock inoculum using the High Pure Viral Nucleic Acid Kit (Roche), per manufacturer protocol. Each sample was isolated individually in chronological order starting with the SIV_{mac}251 inoculum using the High Pure Viral Nucleic Acid Kit (Roche) per manufacturer protocol. vRNA was eluted in 50 μl of elution buffer, aliquoted and immediately stored at -80˚C.

**RT-PCR amplification and library preparation for Illumina deep sequencing**

A unidirectional flow of sample processing was followed for DNase treatment and RT-PCR amplification. RT-PCR, DNase, and Taq control reactions were prepared in a reagent only hood. DNase treatment and RT-PCR amplification were then carried out in a separate RNA/DNA template only hood. vRNA samples were thawed from -80˚C and, immediately prior to RT and PCR amplification, 8 μl of vRNA was DNase treated using amplification grade DNase I followed by a 10 minute incubation with EDTA (Invitrogen, Carlsbad, CA) according to the manufacture’s protocol. Two primer pairs (F1: 5’-GGC CTTCGAATGGCCTAAACAG-3’ and R1: 5’-CCTGCCTTAACCTTAGCTAGC-3’; F2: 5’-TGCACAGGCTTTGGAACAGA-3’ and R2: 5’-ACATCCCCCCTTGTGGAAG-3’) were synthesized with a 5’ amino modifier C6 (Integrated DNA Technologies) to produce two overlapping 2.8kb PCR
amplicon products that spanned the entire SIV env open reading frame. F1 and R1 were used to produce amplicon 1 and F2 and R2 were used to produce amplicon 2. F1 and R1 were initially reported by Bixby et al. [300] as 412Eyu and 413EYd, respectively, and were designed to amplify SIVmac251 viral RNA. 4-5 μl of DNase-treated vRNA was reverse transcribed and PCR amplified using the QIAGEN OneStep RT-PCR kit (Qiagen). Thermal cycling conditions were 45˚C for 120 minutes, 95˚C for 15 minutes followed by 40 cycles of 94˚C for 15 seconds, 50˚C for 30 seconds, and 68˚C for 6 minutes, and a final extension step of 68˚C for 6 minutes. In parallel, 1 μl of DNase-treated vRNA was amplified using the Premix Taq DNA polymerase (Ex Taq version 2.0) (Takara) as a Taq control using the same cycling conditions. An RT-PCR with water instead of template was also performed with each primer pair as a water control in parallel with each RT-PCR reaction. 5 μl of the RT-PCR and Taq reactions were analyzed for each sample on a 1% 1X TAE gel to verify the presence of product in RT-PCR reactions and no product in Taq and water control reactions.

Samples were then PCR purified using the QIAquick PCR purification kit (Qiagen, Germantown, MD) per manufacturer protocol. Samples were analyzed for DNA concentration using a NanoDrop Lite Spectrophotometer (Thermo Scientific, Carlsbad, CA) and approximately 1 ng of each reaction was pooled for shearing, tagmentation, and library preparation (Illumina, San Diego, CA). Each sample was indexed using the Nextera Index Primer Kit (Illumina, San Diego, CA) and sequenced in a single run on an Illumina Miseq using a 300-cycle MiSeq v2 kit (Illumina, San Diego, CA).

**Assembly of deep sequence read data to produce de novo consensus sequences**

For animal Mm 198-08 we captured deep sequence data spanning the entire env open reading frame from 11 longitudinal time points (weeks 2, 3, 4, 6, 8, 10, 14, 18, 23, 25, and 29). For animal Mm 10-10 we captured deep sequence data from 10 longitudinal time points (weeks 2, 4, 6, 8, 10, 14, 18, 23, 25, and 29). For animals Mm 156-10 and Mm 174-08 sampling was limited to 4 longitudinal time points. Animal 156-10, which progressed to disease rapidly, was sacrificed early at 15 weeks post infection (p.i.), however, we captured deep sequence data at 3, 7, 11, and 15 weeks post-infection. In contrast, Mm 174-08 had continuously decreasing set viral loads through the course of infection. We
captured deep sequence data from 2, 3, 7 and 13 weeks of infection. Although, we failed to amplify SIV RNA beyond 13 weeks post-infection because of the low copy number of SIV RNA present in plasma samples using our two-amplicon system, we did amplify sequences spanning V1 and V4 from 17, 22, 24, and 28 weeks of infection.

150 base pair (bp) paired-end sequence reads for each sample were assembled de novo to produce consensus sequences using the publicly available VICUNA software program developed by the Broad Viral Genomics Group [301]. This software program was originally developed to assemble HIV-1 deep sequencing read data, and the parameters used here were performed as previously described [291].

**Detection of sequence variants**

To produce variant calls the deep sequence reads were assembled against the de novo consensus sequence for each sample using the V-Phaser software program developed by the Broad Viral Genomics Group [291, 302]. This process involves assembling all the reads to produce the de novo consensus using Mosaik (http://bioinformatics.bc.edu/marthlab/Mosaik) followed by a step that determines true variants from sequence errors based on an empirically determined error rate, as previously described [291]. Codon frequencies of variant calls were then calculated using the V-Profiler software program [291]. Variant calls made against the consensus sequence of it’s own sample constituted sample diversity and variant calls made against the consensus of the SIV\textsubscript{mac}251 stock inoculum sequence constituted sample divergence. Both the de novo assembly and variant calling were performed by the Viral Genomics Group at the Broad Institute.

**Cloning of SIV\textsubscript{mac}316 Env mutants into full-length SIV\textsubscript{mac}239 viral expression plasmids**

SIV\textsubscript{mac}239 and SIV\textsubscript{mac}251 are neutralization resistant to rhesus monoclonal antibodies (Rh mAbs), therefore to test the neutralization sensitivity of Env substitutions observed during in vivo SIV\textsubscript{mac}239 and SIV\textsubscript{mac}251 replication we engineered single substitution Env mutants in the background of neutralization sensitive SIV\textsubscript{mac}316. SIV\textsubscript{mac}316 has 8 single point mutations resulting in 8 substitutions in Env relative to SIV\textsubscript{mac}239 that increase its neutralization sensitivity. Gene-art constructs based on the SIV\textsubscript{mac}316 nucleotide sequence from an SphI site to an NheI site of about ~2.3 kbs were ordered to introduce single
point mutations by site-directed mutagenesis. 316 Env mutant plasmids were sequenced to verify the presence of the desired mutation and the absence of unintended changes and then digested with SphI and NheI and ligated with the p239 3’ half plasmid digested with the same enzymes. The p239 3’ half plasmids containing 316 Env sequence were then digested with SphI and XhoI and ligated with the p239 5’ half plasmid digested with SphI and XhoI to produce full-length p239 viral expression plasmids containing 316 env mutant sequences.

**Transfection and production of virus stocks**

Virus stocks were produced by transient transfection of 5.5 x 10^6 HEK293T/17 cells with 10 μg of plasmid DNA in D3 media. Four hours post-transfection media was removed and replaced with fresh D10 media. Three days posttransfection the supernatant was harvested and centrifuged at 3000 rpm for 5 minutes to remove cellular debris. Virus containing supernatant stocks were aliquoted at 1 ml or 500 μl at immediately stored at -80°C. The virus content of the supernatant stocks was determined by an SIV p27 Antigen Capture Assay Kit (Advanced Bioscience Laboratories, Rockville, MD), following the manufacturer’s protocol.

**Virus Infectivity Assay**

Infections were performed in 96-well plates. To determine the virus infectivity of each virus stock, 100 μl of virus-containing supernatant-stock was serially diluted in 100 μl of R10 media (1:2 dilutional series). 100 μl of R10 media containing 1 x 10^3 TZM-bl cells, which contain a stably integrated Tat-responsive LTR promoter driving beta-galactosidase and luciferase activity, was then added to wells containing 100 μl of each diluted stock, bringing the final volume to 200 μl. Plates were then incubated at 37°C at 5% CO2. 72 hours post-infection the level of luciferase activity in each well was measured using the britelite plus Reporter Gene Assay System kit (PerkinElmer, Waltham, MA). 100 μl of supernatant was removed the assay plate and 100 μl of reconstituted britelite reagent (luciferase substrate) was added to each well. Cells were lysed by mixing and 150 μl of lysed cell containing supernatant was added to the test plate to measure luminescence. Luminescence was measured using a Victor X5 plate reader (PerkinElmer, Waltham, MA).
**Virus neutralization assays**

Neutralization assays were performed to measure the neutralizing activity of rhesus mAbs against SIV strains SIV\textsubscript{mac}239 and SIV\textsubscript{mac}316 and to measure the effect of substitutions in SIV gp120 on neutralization sensitivity against Rh mAbs in the context of SIV\textsubscript{mac}316. Neutralization assays were performed using a C8166-derived CD4\textsuperscript{+} T cell line transduced with an alkaline phosphatase reporter gene under the transcriptional control of the SIV LTR promoter, as previously described [296]. Neutralization was determined as the level of inhibition of secreted alkaline phosphatase (SEAP) activity over serial four-fold dilutions of mAb. Different amounts of virus input were used in experiments based on virus infectivity of SIV\textsubscript{mac}239 and SIV\textsubscript{mac}316. Virus equivalent to 1.5 ng p27 was used for experiments with SIV\textsubscript{mac}239 WT and 10 ng p27 was used for experiments with SIV\textsubscript{mac}316 or single substitution SIV\textsubscript{mac}316 mutants.

Neutralization assays were performed in 96 well plates using 5,000 C8166-SEAP cells per well. Cells in column 1 were incubated with no virus providing a background level of SEAP activity. Cells in column 2 received virus but no mAb providing a maximal level of SEAP activity. First, an appropriate volume of R10 media was added to all wells. Next, column 3 received mAb at a starting concentration of 10 μg/μl followed by nine four-fold serial dilutions across the plate to column 12. Virus was then added to columns 2-12 in a volume of 50 μl and the plates were incubated for 1 hour at 37°C at 5% CO\textsubscript{2} levels. After 1 hour 5,000 SEAP cells were added to every well in a volume of 75 μl bringing the final volume in each well to 200 μl. The level of SEAP activity in the supernatant was measured after 72 hours for wells that received SIV\textsubscript{mac}239. SEAP activity was measured after 5 days for wells that received SIV\textsubscript{mac}316 or 316 Env mutant viruses. SEAP activity was measured using the Phospha-Light SEAP Detection Kit (Applied Biosystems, Foster City, CA), according to manufacturer protocol.

Neutralization assays were also performed using soluble CD4-183 (sCD4-183), which contains the first 183 residues of CD4 consisting of the two N-terminal domains of CD4. The neutralization titer of sCD4-183 was measured against SIV\textsubscript{mac}239 and single substitution SIV\textsubscript{mac}239 Env gp120 mutant
viruses. 1.5 ng p27 of virus input was used for SIV<sub>mac</sub>239 WT and 239 Env mutants. The assays were performed as described above, except the starting concentration used for sCD4-183 was 25 µg/µl.

**ADCC assay**

The target cells for the ADCC assay are CEM.NKR-CCR5 CD4<sup>+</sup> T cells transduced with the firefly luciferase gene under the control of a SIV LTR promoter. The effector cells are KHYG-1 human NK cells that were transduced with cDNA that express rhesus CD16. Both cell lines were used to carry out ADCC assays, as previously described [299]. Target cells were infected by spinoculation with 200 ng of SIV<sub>mac</sub>239, single substitution Env mutants, or 239 Env mutants consisting of multiple amino acid deletion mutants within the V4 loop region. Spinoculation was carried out at 12000 X g for 2 hours. The viral inoculum was then removed and the infected targets were resuspended in fresh R10 media and incubated for 96 hours at 37°C at 5% CO<sub>2</sub> levels. On the day of the assay the infected target cells were washed 3 times with R10 media. Then NK cells, infected target cells, and uninfected target cells were counted and target cells were mixed with NK effector cells at a ratio of 1:10 with a total of 10<sup>4</sup> target cells to 10<sup>5</sup> NK effector cells per well in a 96-well plate in a final volume of 200 µl. Uninfected target cells mixed with NK cells, but not plasma served as background and defined 0% relative light units (RLU). Infected target cells mixed with NK effector cells, but not plasma defined 100% RLU for each SIV strain. To measure the ADCC activity in plasma samples from our cohort of SIV-infected macaques we prepared a serial two-fold dilution of terminal plasma samples for the four animals starting at a 1:32 dilution. 50 µl of serially diluted plasma was added to wells of infected targets with NK effector cells in a total volume of 200 µl. Plates were then incubated for 8 hours at 37°C at 5% CO<sub>2</sub>. 8 hours post-infection the level of luciferase activity in each well was measured using the britelite plus Reporter Gene Assay System kit (PerkinElmer, Waltham, MA). After the 8-hour incubation, 150 µl from each well was resuspended and mixed with 50 µl of luciferase substrate. The level of luciferase activity in each well was determined using a plate reader.

Non-Env specific antibodies that recognize antigens on the surface of the human target cells can interfere with Env-specific ADCC activity [299]. To deplete plasma samples of antibodies that bind to
human antigens $10^7$ CEM.NKR target cells were pelleted and resuspended with each plasma sample. Plasma was incubated with target cells at room temperature for twenty minutes before being centrifuged, removed, and used to resuspend another pellet of target cells. Six rounds of depletions were carried out with each plasma sample before being used in the assay.

**Recombinant soluble gp140 protein production**

A codon-optimized expression vector for soluble, recombinant SIV \text{mac}239 Env gp140 was described previously [303]. gp140 contains the gp120 subunit and the N-terminal ectodomain of gp41, but is cleaved prior to the TM region. SIV_{\text{mac}}239 gp140 proteins were found to express as stable soluble trimer proteins, similar to membrane expressed Env trimers [303]. Soluble recombinant expression-optimized SIV_{\text{mac}}239 gp140 WT and five gp140 single point mutants were synthesized by transfection of HEK293T/17 cells with our WT and mutant SIV_{\text{mac}}239 gp140 expression vectors. Cellular supernatants were harvested 2 days post transfection and 5 days post transfection. Supernatants were pooled and centrifuged to remove cellular debris, passed through a 0.45 µm filter, and purified using Galanthus nivalis Lectin-Agarose beads. To confirm expression of our WT and mutant SIV_{\text{mac}}239 gp140 proteins we separated the purified non-biotinylated proteins on a denaturing SDS-PAGE gel in reducing conditions and probed for the presence of soluble gp140 through coomassie stain and by western blot using rhesus (Rh) mAb 3.11H which recognizes a linear epitope in the V3 loop of SIV gp120 (data not shown).

**Rhesus monoclonal antibody production from Epstein Barr virus-immortalized rhesus B cell lines**

Rhesus monoclonal antibodies (Rh mAbs) specific for the gp120 subunit of SIV Env were isolated from serum-free cellular supernatants of immortalized hybridoma B-lymphocyte cell lines (BLCLs) to perform ELISA-based binding assays against WT and mutant SIV_{\text{mac}}239 gp140 proteins. To isolate the Rh mAbs, BLCLs were passaged over several weeks to produce 0.5 to 1 liter of supernatant. The supernatants were harvested, centrifuged twice to reduce cellular debris, passed through a 0.45 µm filter, and purified with a Protein A/Protein G resin mix. The concentration of the mAbs was determined by absorbance at 280 nm using a spectrophotometer.
**SIVmac239 gp140 binding by Rh mAbs by ELISA**

Wells of a 96-well plate were coated with 50 ngs SIVmac239 WT or mutant gp140 proteins overnight at 4°C. The following morning the plates were washed and blocked with 1X PBS with 5% FBS for 2 hours at 37°C at 5% CO₂. Plates were washed again before adding a serial six-fold dilution of the mAb being tested for binding at a starting concentration of 10 µg/ul. The plates were then incubated for one hour at 37°C at 5% CO₂. The plates were washed again and a HRP-conjugated anti-rhesus IgG was added as a secondary antibody. The plates were incubated again at 37°C for one hour. Detection was measured with the addition of TMB substrate followed by stop solution and absorbance was read at 450 nm.
2.5 Results

2.5.1 Antibody responses to SIV Envelope in cohort of SIV-infected rhesus macaques.

To understand SIV Env sequence evolution in the face of Env-specific antibody responses, we first determined the presence of antibody responses in a cohort of four rhesus macaques experimentally infected by repeated low dose mucosal exposure to uncloned SIVmac251 (Table 1). The animals were genotyped for TRIM5 and for MHC class I alleles (Table 1). SIV viral RNA loads measured from longitudinal plasma samples verified all four animals were successfully infected, with acute peak viral replication ranging from $5.5 \times 10^6$ to $2.3 \times 10^8$ copies/ml (Table 1;Figure 6). Animal Mm 156-10 retained viral loads above 10 million copies/ml through 15 weeks of infection and was euthanized early due to the rapid development of AIDS-related symptoms. The remaining 3 animals showed typical post-acute viral loads characteristic of SIVmac-infected macaques (Figure 6). To determine whether each animal mounted an antibody response specific for Env, we screened for binding antibodies of longitudinal plasma samples to soluble SIVmac239 Env gp140 by ELISA (Figure 6). We found that animals Mm10-10, 198-08, and 174-08 did have antibody responses specific for Env that increased in titer over time, but we did not detect an antibody response to Env in Mm 156-10 at any time-point tested. We further tested for neutralization of a lab-adapted stock of SIVmac251 (Figure 6). We found that plasma from all animals except for Mm 156-10 could neutralize lab adapted SIVmac251. Therefore, we had 3 animals with typical viral load courses and detectable antibody responses and one rapid progressor animal that did not mount a detectable antibody response to Env. We next used this small cohort of animals to track the evolution of the SIV env gene in the environment of a sustained Env-specific antibody response.
Figure 6. Plasma viral RNA load and antibody-based immune responses following repeated low-dose SIVmac251 mucosal challenge.

A. Viral loads of 4 rhesus macaques infected by repeated low dose challenge with SIVmac251 were measured 1, 2, 3, 4, 6, 8, 10, 14, 18, 23, 25, and 29 weeks from initial challenge by real time RT-PCR of plasma viral RNA. All four animals were positively infected showing peak of viral replication 2 weeks from the last week viral RNA was not detected. Viral loads were synchronized to the last week in which vRNA was not detectable in each infected animal. B. Env-specific antibody responses were measured by ELISA using recombinant soluble SIVmac239 gp140. The level of binding in plasma from each animal was tested by serial 4-fold dilution in duplicate. Antibody titer was calculated as the inverse of the lowest dilution with a value at least two-fold above background. C. Neutralizing antibody titers were measured from plasma against lab-adapted SIVmac251. Plasma was tested over eight serial two-fold dilutions in duplicate to determine the concentration of plasma which reduce infection by 50% or 50% IC. The 50% IC was reported for every time point tested for each animal in the cohort.

2.5.2 High-resolution deep sequencing of the SIV env gene during replication in rhesus macaques

To investigate SIV env sequence evolution during the earliest stages of SIV infection, we used Illumina deep sequencing of RT-PCR amplified SIV RNA extracted from longitudinal plasma samples of each animal. Sequence data were produced using an overlapping two-amplicon system that targeted the entire env coding region (~4 kilobases) (Figure 7). Each RT-PCR amplified SIV RNA sample was uniquely indexed and pooled before being sequenced to produce 150 base-pair paired-end reads. Reads from each longitudinal sample and the stock inoculum sample were assembled de novo to produce
Two primer sets were used to amplify viral RNA spanning the \textit{env} open reading frame. F1 (412EYu) binds to positions 6127-6147 in \textit{vif/vpx} of the SIV$_{mac}$239 genome and R1 (413EYd) binds to positions 8898-9017. F2 binds to positions 7328-7347 in \textit{env} and R2 binds to positions 10,136-10,153 in the 3' LTR. Each amplicon is approximately 2.8-2.9 kilobase (kb) pairs and overlap by 1.7 kb to produce 4 kbs of coverage spanning the 3' half of the SIV$_{mac}$239 genome. All primers were ordered with a 5' amino modifier C6 (IDT). F1 and F2 were initially reported by Bixby \textit{et al.} [300].

Figure 7. Location of two primer set to produce amplicons spanning the \textit{env}.
Two primer sets were used to amplify viral RNA spanning the \textit{env} open reading frame. F1 (412EYu) binds to positions 6127-6147 in \textit{vif/vpx} of the SIV$_{mac}$239 genome and R1 (413EYd) binds to positions 8898-9017. F2 binds to positions 7328-7347 in \textit{env} and R2 binds to positions 10,136-10,153 in the 3' LTR. Each amplicon is approximately 2.8-2.9 kilobase (kb) pairs and overlap by 1.7 kb to produce 4 kbs of coverage spanning the 3' half of the SIV$_{mac}$239 genome. All primers were ordered with a 5' amino modifier C6 (IDT). F1 and F2 were initially reported by Bixby \textit{et al.} [300].

2.5.3 Diversity in stock of uncloned SIV$_{mac}$251 swarm captured by Illumina deep sequencing

In order to capture the dynamics of mucosal SIV transmission, we performed deep sequencing of the SIV$_{mac}$251 stock inoculum as well as longitudinal samples spanning acute and early infection. We first determined diversity within each sample (including the stock). Variant calls were made by assembling the reads for each sample/timepoint to the consensus for that sample/timepoint. We defined Env diversity as the frequency of codons present at each site in Env in addition to the majority codon. For the SIV$_{mac}$251 stock, we determined that the mean diversity across all sites in Env was 0.555 for non-synonymous (NS) diversity and 0.416 for synonymous (S) diversity (Figure 9). The mean NS diversity for gp120 and gp41 was 0.425 and 0.751, respectively, while the S diversity for gp120 and gp41 was 0.353 and 0.511, respectively. Focusing on the variable loop regions (V1-V5) of gp120, the mean NS
Figure 8. Read coverage of SIVmac251 env gene.

Deep sequencing reads produced by Illumina deep sequencing spanning the env gene from a swarm stock of SIVmac251 and longitudinal plasma samples from (B) Mm 174-08, (C) Mm 10-10, (D) Mm 156-10, and (E) Mm 198-08. The reads were produced from two overlapping amplicons spanning the env open reading frame.

Diversity ranged from 0.022 to 0.752 for V2 through V5, but, strikingly, was 2.477 for the V1 loop. Positions K120, S132, and S139 (SIVmac239 numbering) in V1 all had secondary residues present with frequencies above 14%. A large degree of NS variation in gp41 was present in the cytoplasmic tail with eight sites with secondary residues present at frequencies above 10% and three of those sites (F732, S785,
and Q831) having secondary residues with frequencies above 42%. The degree of genetic variation we observed in our stock matched other estimates of diversity of uncloned SIV<sub>mac</sub>251 measured through single genome amplification [304].

![Graph showing NS and S diversity in SIV<sub>mac</sub>251](image)

**Figure 9.** Non-synonymous and synonymous diversity present in an uncloned stock of SIV<sub>mac</sub>251.

The level of NS and S diversity was determined for every site in Env. NS diversity was determined as the sum of the frequency of any codon(s) other than the majority codon in the consensus sequence of the SIV<sub>mac</sub>251 stock that did not code for the same amino acid (blue dots). S diversity was determined as the frequency of any codon other than the majority codon in the consensus sequence of the SIV<sub>mac</sub>251 stock that coded for the same amino acid as the majority codon (red dots). The red boxes highlight the two regions of the highest variation in env: the V1 region in gp120 and the cytoplasmic tail in gp41.

### 2.5.4 Loss of diversity and divergence from the stock reflects a transmission bottleneck during mucosal transmission of uncloned SIV<sub>mac</sub>251

We saw a uniform loss of diversity at the first time point sampled in each animal (either 2 or 3 weeks post-infection) relative to the diversity observed in the stock inoculum, indicative of a transmission bottleneck event (Figure 10). We next examined divergence of timepoints from the stock; we defined divergence as a change in the consensus residue at each position relative to the stock consensus.
Figure 10. Diversity of env gene in virus populations at acute SIV infection of (~2-3 weeks).
The level of NS and S diversity was determined for every site in Env. NS diversity was determined as the sum of the frequency of any codon(s) other than the majority codon in the consensus sequence of the sample that did not code for the same amino acid (blue dots). S diversity was determined as the frequency of any codon other than the majority codon in the consensus sequence of the sample that coded for the same amino acid as the majority codon (red dots). As the majority codon had to be present at 50% or more, viral diversity was limited to 50%. NS: non-synonymous; S: synonymous.

Comparison of the stock consensus to the consensus sequences at the earliest sampled time point in each animal revealed the presence of a 3 amino-acid insertion deletion (Pro-Thr-Ala) in the hypervariable region of V1 present in the stock that only transmitted into 1 of 4 animals, Mm 174-08. The insertion was retained 3 weeks into infection and subsequently lost by 7 weeks post-infection. The same 3 amino acid insertion emerged in Mm 10-10 at 6 weeks p.i. but then was subsequently lost in the population by 8 weeks p.i. Similarly, a 2 amino acid repeat in the V4 loop (LTLT) was not transmitted to any animal, resulting in the loss of two amino acids in the V4 loop in each animal after transmission. The two amino
acids did not re-emerge in any of the animals at any of the time points we sampled. Therefore, we observed a preference for SIV variants with shorter V1 and V4 loops during mucosal transmission into each animal.

We identified several sites of divergence at the earliest sampled time point in each animal clustering in the V1 loop of gp120 and the C-terminal region of the cytoplasmic tail of gp41 (Figure 11).

Figure 11. Sites of divergence from the SIVmac251 stock inoculum at two or three weeks post-infection (acute infection) in four rhesus macaques challenged by repeated low dose challenge.

NS sites are reported in blue and S sites are reported in red. Divergence was determined as the frequency of a majority codon (frequency of 50% or more) in the sampled time-point that differed from the majority codon of the SIVmac251 stock inoculum.

In many instances the residues that diverged were present at low frequency in the stock inoculum, suggesting transmission of minor or rare variants from the viral population present in the stock inoculum.

2.5.5 Reestablishment of diversity from a homogeneous population during acute infection.
For each sample we calculated the NS and S diversity at each site in Env over acute, post-acute, and early chronic stages of infection (Figure 12). We observed a reestablishment of diversity in Mm 10-

**Figure 12. Loss of viral diversity during transmission of SIVmac251 before emergence of viral diversity during early SIV infection in rhesus macaques.**

The level of NS and S diversity was determined for every site in Env. NS diversity was determined as the sum of the frequency of any codon(s) other then the majority codon in the consensus sequence of that sample that did not code for the same amino acid (red) and S diversity was determined as the sum of the frequency of any codon(s) other then the majority codon in the consensus sequence of that sample that did not code for a different amino acid.
10 and Mm 198-08 detectable by 14 weeks p.i. At the latest sampled timepoint (29 weeks p.i.) for both Mm 10-10 and Mm 198-08 we observed a high level of Env diversity that matched the stock inoculum. In contrast, diversity in Mm 156-10 and 174-08 remained low. At the latest sampled timepoint for Mm 156-10 (15 weeks p.i.) we found only eight sites in gp120 with NS diversity over 10% and only a single site in the HR2 region of gp41 with diversity over 10%. Similarly, only a single site of S diversity in the HR1 region of gp41 rose above 10% (Figure 12). Interestingly, we found only three sites of NS diversity and two sites of S diversity that rose above 10% in Mm 174-08, which had detectable antibody responses, evidence that at 13 weeks p.i. env diversity was still low relative to the level of diversity present in Mm 198-08 and Mm 10-10 (Figure 12).

**Evolution of the V1 and V4 variable regions during early infection.**

Overall, while we found variation scattered throughout Env, variation was significantly higher in the V1 and V4 variable domains. Variation in these domains and a potential role in escape from neutralizing antibody has been reported by others [161, 162, 176, 246, 247]. We therefore made a closer examination of the evolution of these two domains in the viral population throughout early infection.

*Evolution of the SIVmac251 Env V1 loop during early infection.* Comparing the mean diversity of each variable region to the total mean diversity of Env, we observed the highest diversity emerge in V1 and V4 in animals Mm 198-08 and Mm 10-10. We found a blip of V1 codon diversity occurred at 6 weeks p.i. in animals Mm 198-08 and Mm 10-10 that was followed by a period of low level sequence variation until diversity in V1 began to increase again by week 14 in Mm 198-08 and weeks 18 and 23 in Mm 10-10 (Figure 13).

We did not observe the same level of Env diversity in Mm 156-10 at any time-point up to 15 weeks in V1. This animal maintained viral loads above 6 million copies/ml through 15 weeks of infection. Similar to Mm 156-10, Mm 174-08 did not evolve Env diversity, either within variable regions or across Env, up to weeks 13 post-infection (Figure 13), yet this animal did mount a detectable antibody response similar to Mm 198-08 and Mm 10-10 (Figure 6).
Figure 13. Evolution of sequence diversity in SIV Env V1 loop in Mm 10-10, Mm 198-08, Mm 174-08, and Mm 156-10. The mean diversity of all sites in Env (black lines) and just the V1 loop region (green lines) plotted over 29 weeks of SIV replication in Mm 10-10 and 198-08, 13 weeks in Mm 174-08, and 15 weeks in Mm 156-10.

We found changes were tightly clustered within a patch of approximately 12 codons in V1 in both Mm 198-08 and 10-10; similar patterns have been observed in SIV Env V1 sequences isolated from rhesus macaques, pig-tailed macaques and sooty mangabey monkeys [162, 177, 201, 246, 247, 258, 305, 306]. In Mm 10-10 V1 we found a predominance of variation at serine, threonine, and alanine sites, specifically S127, S132, T135, T136, A138, and S139, in addition to one site outside of this hypervariable region, K120 (Figure 14; positions are identified as the consensus residue in the stock inoculum but based on SIVmac239 numbering). The blip in V1 diversity we observed in Mm 10-10 was the result of variation at four sites: K120, S127, S132, and S139. Each of these four sites had two or more residues present in
Figure 14. Evolution of sequence diversity in SIV Env V1 loop in Mm 10-10.
The frequency of codons were determined from deep sequence reads of each sample assembled to the de novo consensus sequence of the same sample using the vphaser software program (Broad Research Institute). The 7 sites within the hypervariable region of V1 is plotted over the first 29 weeks of in vivo viral replication and are representative of sites with the highest codon variation.

The viral population of the stock inoculum with only one residue being present at 100% frequency at each site 2 weeks into infection, the peak of viral load, except for position 139, which possessed S139P as a minor variant at a frequency of 0.85%. At six weeks p.i lysine rose to a frequency of 33.64% at position 120 through a single nucleotide change (AGA → AAA) and at position 132 threonine rose to a frequency of 40.3%, also through a single nucleotide change (TCA → ACA). Variation was less pronounced at positions 127 and 139 with serine emerging at position 127 to a frequency of 15.93% and proline emerging to a frequency of 9.94%, with subtle differences in the dynamics of variation at each site. For
instance, the change at position 127 required two nucleotide changes (ATA → TCA). Leucine (TTA) was present in the inoculum at position 127, thus it is possible that single nucleotide changes from ATA to TTA then to TCA allowed for the emergence of serine, however, leucine was not detected at 2 or 3 weeks into infection. Additionally, at position 139 proline was present 2 weeks into infection at a low frequency of 0.85% before peaking at 9.94% at six weeks and persisting at a low level until 14 weeks p.i. Interestingly, at the three other sites (120, 127, and 132) the ancestral residue returned to 100% frequency by the next sampled time point, 8 weeks p.i.

Variation at the remaining sites followed a similar pattern of change, with all sites having a single dominant residue present at 100% until the emergence of variants beginning at either 18 or 23 weeks into infection. Variation at positions 120, 132, 135, 136, and 138 resulted in the newly emerged codon variant replacing the ancestral codon and becoming dominant in the viral population.

The dynamics of V1 variation in Mm 198-08 differed from Mm 10-10 however, variation clustered within the same hypervariable regions and at many of the same sites (Figure 15). The most striking difference was the level of change that resulted in complete replacement of the residue from the previous time point occurring at positions 120, 127, 132, 135, and 139. This phenomena was most striking at position 135 with all possible codons with _TC nucleotides at the 2nd and 3rd position being present (i.e., GCA, TCA, ACA, and CCA) over the course of infection. In fact, in contrast to the dynamics observed in Mm 10-10 the level of diversity in the hypervariable region of V1 in Mm 198-08 continued to increase from 14 weeks until the end of the sampled time period (29 weeks p.i.)(Figure 13).

Deep sequencing of well-spaced longitudinal samples allowed us to capture the dynamics in V1 length variation in the viral population of Mm 198-08, particularly from 18 to 29 weeks p.i. V1 length variation in Mm 198-08 began between weeks 14 and 18 and continued through 29 weeks p.i. Between 14 and 18 weeks p.i. a 9 base pair deletion occurred at the viral population level changing the consensus
Figure 15. Evolution of sequence diversity in SIV Env V1 loop in Mm 198-08.

The frequency of codons were determined from deep sequence reads of each sample assembled to the de novo consensus sequence of the same sample using the vphaser software program (Broad Research Institute). The 7 sites within the hypervariable region of V1 is plotted over the first 29 weeks of in vivo viral replication and are representative of sites with the highest codon variation.

V1 sequence from 131-ASTTTTAPAK-141 to 131-TAAPTAMS-138 (Figure 16). This was then followed by the complete replacement of 131-TAAPTAMS-138 with 131-ASTTTASTTATTAPAK-147 by week 23, resulting in a six amino acid increase in length relative to week 14. The insertion consisted of codons GCA, ACA, and TCA, resulting in the insertion of alanine, serine, and threonine amino acid residues. This pattern of V1 length insertions continued at weeks 25 increasing the length of the consensus followed by a retraction in V1 length at 29 weeks p.i.

The rapid emergence of serine, threonine, and alanine insertions and substitutions in both Mm V1 loop hypervariable region

70
| SIVmac239 | CNKSETDRWGLTKSITTTAS−−−−−−−−−−−−TTSTTA−SAKIDMVNETSSC |
| SIVmac251 | CNKSETDKWLTKSITTTAP−−−−−−−−−−−−TTTTTA−PAKIDMVNETSSC |
| wk 2     | CNKSETDKWLTKSLTTTAP−−−−−−−−−−−−TTAPAA−SAKIDMVNETSSC |
| wk 3     | CNKSETDKWLTKSLTTTAP−−−−−−−−−−−−TTAPAA−S−KIDMVNETSSC |
| wk 4     | CNKSETDKWLTKSLTTTAP−−−−−−−−−−−−TTAPAA−S−KIDMVNETSSC |
| wk 6     | CNKSETDKWLTKSLTTTAP−−−−−−−−−−−−TTAPAA−S−KIDMVNETSSC |
| wk 8     | CNKSETDKWLTKSITTTAS−−−−−−−−−−−−TTTTTA−PAKIDMVNETSSC |
| wk 10    | CNKSETDKWLTKSITTTAS−−−−−−−−−−−−TTTTTA−PAKIDMVNETSSC |
| wk 14    | CNKSETDKWLTKSITTTAS−−−−−−−−−−−−TTTTTA−PAKIDMVNETSSC |
| wk 18    | CNKSETDRWGLTKSLTTTT−−−−−−−−−−−−AAAT−AMS−IDMVNETSPC |
| wk 23    | CNKSETDKWLTKSITTTAS−−−−−−−−−−−−TTTASTTATT−PAKIDMVNETSSC |
| wk 25    | CNKSETDKWLTKSITTTAS−−−−−−−−−−−−TTTASTTATT−PAKIDMVNETSSC |
| wk 29    | CNKSETDKWLTKSITTTAS−−−−−−−−−−−−TTTASTTATT−PAKIDMVNETSSC |

Figure 16. Consensus sequences of the Env V1 loop region in Mm 198-08 over early SIV infection.

The consensus sequences were produced from Illumina deep sequence reads of longitudinal plasma viral RNA samples and viral RNA isolated from the stock inoculum using the v-profiler software program (Broad Research Institute). Dashes represent insertions or deletions relative to the consensus of the SIVmac251 stock V1 sequence. For reference the V1 loop sequence of SIVmac239 is included in the top row. Positions 113 and 154 are in reference to the conserved cysteines that form the disulfide bond producing the V1 loop and are based on SIVmac239 sequence.

198-08 and Mm 10-10 is concurrent with sequence changes resulting in O-linked glycan modifications, which are predicted to occur along stretches of serine and threonine residues [175, 260].

Interestingly, threonine at position 136 is coded for by ACG unlike every other threonine in the V1 loop, which is coded for by ACA. This ACG codon was present in the SIVmac251 stock inoculum and was maintained in three of four animals in our cohort (ACG was replaced by AAT by 29 weeks p.i. in Mm 10-10) and is observed in other SIVmac variants including SIVmac239. Strikingly, during the large fluctuations in V1 length observed in Mm 198-08 this ACG codon was never lost or replaced and no length variation was observed beyond this ACG codon. Although speculative, perhaps the ACA/TCA/GCA codon rich region is more prone to slippage by RT during reverse transcription, providing a mechanism for rapid adaptation in the V1 domain. If true, perhaps the ACG codon may provide a mechanism in which RT is able to maintain subsequent read through that maintains the proper reading frame for env, however, this would need to be verified through further investigation.

Evolution of the SIVmac251 Env V4 loop during early infection includes frequent insertion/deletions and substitutions. The V4 loop of Env evolved a similar level of sequence variation as the V1 loop during viral replication in Mm 198-08 and Mm 10-10, but again we did not find similar levels of variability in the viral populations of Mm 156-10 or Mm 174-08 (Figure 17). The mean diversity
of V4 began to rise at 14 weeks p.i. in Mm 10-10 and as early as 10 weeks p.i. in Mm 198-08, but we found no increase in diversity in animals Mm 174-08 and Mm 156-10 during the same time period (Figure 17).

Variations that included substitutions and large in-frame deletions were tightly clustered within a region spanning approximately 9 amino acids from positions 415 to 424. Substitutions were

![Figure 17. Evolution of sequence diversity in SIV Env V4 loop in Mm 10-10, Mm 198-08, Mm 174-08, and Mm 156-10. The mean diversity of Env and the V1 loop region is plotted over 29 weeks of SIV replication in Mm 10-10 and Mm 198-08, 13 weeks in Mm 174-08, and 15 weeks in Mm 156-10.](image-url)
predominantly found at three sites in the V4 loops of Mm 198-08 and Mm 10-10: these were at D415, P421, and R424 (Figure 18). D415N introduces a new N-linked glycosylation motif in the V4

**Figure 18. Evolution of sequence diversity in SIV Env V4 loop.**
The frequency of codons were determined from deep sequence reads of each sample assembled to the de novo consensus sequence of the same sample using the vphaser software program (Broad Research Institute). A-D. Three sites within V4 is plotted over the first 29 weeks of in vivo viral replication of Mm 10-10 (A) and Mm 198-08 (B), over the first 15 weeks of in vivo viral replication in Mm 156-10 (C), and over the first 13 weeks of in vivo viral replication of Mm 174-08 (D), and are representative of sites with the highest codon variation in the V4 loop.
loop. Interestingly, SIV\textsubscript{mac}239 possess an asparagine at position 415, but an alanine at position 417. Thus, a common adaptation in SIV\textsubscript{mac}239 Env is A417T, which, similar to the D415N substitution in SIV\textsubscript{mac}251, introduces an N-linked glycosylation motif [162, 246]. D415N emerged by week 18 in Mm 10-10 at a frequency of 6.58% before peaking at 48.26% by week 25 followed by a sharp decrease to 13.32% at week 29 p.i. In Mm 198-08 D415N emerged by week 10 p.i. and increased to 100% frequency by week 23. D415N remained fixed at 100% at 25 and 29 weeks p.i.

Position P421 experienced some of the highest variation within a single given site in gp120 (Figure 18). Variation at P421 first emerged with a P421S substitution by 14 weeks in Mm 10-10 and 10 weeks in Mm 198-08. Strikingly, in Mm 10-10, we observed the emergence of three additional variants in parallel at 18 weeks p.i. – by this time point, the P421S, P421T, P421L and P421Q variants were all present at 6% or higher reducing the level of P421 to 67.07%. The same variants were maintained up through 23 weeks p.i., but by 25 weeks only P421L and P421Q remained at frequencies of 17.66 and 10.98, respectively. These two variants were then replaced by P421T, which re-emerged to become the dominant residue by 29 weeks p.i. at 82.14% Although variation at P421 in Mm 198-08 was less dynamic than Mm 10-10 with only P421S and P421L rising above a frequency of 5%, each of these variants was observed in Mm 198-08 as well. Interestingly, P421S was observed in Mm 156-10 at a frequency of 16.27% at 11 weeks p.i.

SIV\textsubscript{mac}251 possesses a string of 3 arginine residues at position 424, 426, and 428. Variation was observed at all arginines, but only R424 had variation that rose above 5%. R424Q was observed in both Mm 198-08 and Mm 10-10 (Figure 18). In Mm 198-08 R424 was lost from the viral population at 18 weeks p.i. and then again at 29 weeks p.i. (Figure 18). In fact residues \textsuperscript{425}ERHR\textsuperscript{426} in the V4 loop were not present in the consensus sequence in the V4 loop at 18 weeks p.i. and residues \textsuperscript{424}RH\textsuperscript{425} were not present in the consensus sequence in the V4 loop at 29 weeks p.i.

We detected the presence of deletions in reads spanning the V4 loop, but deletions were only detectable in the V-Phaser output files if the deletion reached a high enough frequency that it was no longer included in the consensus sequence, such as for position 424 in Mm 198-08. Large deletion events
were evident in the reads spanning the V4 loop in Mm 198-08 and Mm 10-10. Therefore we extracted reads spanning the V4 loop from 10, 14, 18, 23, 25, and 29 weeks p.i. for Mm 198-08 and Mm 10-10, aligned the reads to the consensus of the SIVmac251 stock inoculum, and then tallied the frequency of reads that possessed in-frame deletions relative to the total number of reads aligned to the consensus (Figure 19 and Figure 20).

We found reads spanning the V4 loop with deletions in Mm 10-10 were present at 10 weeks p.i., but remained at a frequency below 2% until 18 weeks p.i., when deletion $\Delta^{415}$DLTTQR$^{420}$ was detectable at 2.8% and increased to 35.9% by 23 weeks p.i. (Figure 19). By 25 weeks p.i. two additional deletion variants were detectable: $\Delta^{423}$ERHR$^{426}$ at 16.55% and $\Delta^{421}$PKERHRR$^{427}$ at 9.24%, with $\Delta^{415}$DLTTQR$^{420}$ still present at 13.21%. By 29 weeks p.i. the frequency of $\Delta$ERHR and $\Delta$DLTTQR were reduced to below 5%. $\Delta$PKERHRR remained at a similar frequency of 8.32%, but $\Delta$TQRPKERH rose to a frequency of 45.58%. $\Delta$TQRPKERH was present at 23 weeks p.i. but not observed at 25 weeks p.i. before becoming the dominant deletion variant at 29 weeks p.i., thus either this deletion variant decreased to a level below the threshold of detection or emerged twice during infection. Interestingly, all three deletion variants either flanked or spanned the P421 site in V4, providing strong evidence of selection continuously targeting the V4 loop at or around P421 resulting in replacement of P421 with other residues or deletions at or near P421. We observed a similar pattern of deletions in the V4 loop in Mm 198-08, however, the emergence of deletions occurred much earlier with deletion variants rising to a high frequency by 14 weeks p.i. (Figure 20). Three deletion variants were detectable at 14 weeks: $\Delta$KERH, $\Delta$DLTTQR, and $\Delta$ERHR. $\Delta$ERHR was observed in Mm 10-10 as well. However, unlike in Mm 10-10, $\Delta$ERHR rose to dominate the viral population in Mm 198-08, staying above a level of 65% from 18 weeks p.i. through to 29 weeks p.i. and rising as high as 83.7% at 18 weeks p.i. Deletions spanning P421 in V4 do result in loss of recognition to antibodies that recognize the V4 loop region [249]. The emergence of $\Delta$ERHR in two animals in our cohort, the rise to high frequency early in infection, and persistence of this deletion variant over several weeks of replication, to high levels within Mm 198-08, provides strong evidence that host Env-specific antibodies continuously target the V4 loop. Mm 198-08
Figure 19. Evolution of sequence diversity in SIV Env V4 loop observed in Mm 10-10.

A. The presence of multiple amino acid deletions are aligned relative to the V4 loop sequence of SIVmac239 and the consensus of the SIVmac251 stock inoculum. In-frame multiple amino acid deletions flank or overlap the P421 site (highlighted in red in the 239 and 251 V4 reference sequences). B. The relative frequency of deletion variants detected in the reads spanning the entire V4 loop are reported at 10, 14, 18, 23, 25, and 29 weeks p.i. Reads that did not posses any in-frame deletions were reported as no deletion (in grey).
Figure 20. Evolution of sequence diversity in SIV Env V4 loop observed in Mm 198-08.
A. The presence of multiple amino acid deletions are aligned relative to the V4 loop sequence of SIVmac239 and the consensus of the SIVmac251 stock inoculum. In-frame multiple amino acid deletions flank the P421 site (highlighted in red in the 239 and 251 V4 reference sequences). B. The relative frequency of deletion variants detected in reads spanning the entire V4 loop are reported at 10, 14, 18, 23, 25, and 29 weeks p.i. Reads that did not possess any in-frame deletions were reported as no deletion (in grey).
maintained viral loads above $10^5$ copies/ml during this period of early infection suggesting that Env has evolved mechanisms to allow for multiple amino acid deletions within variable loop regions that do not have an effect on the ability of Env to maintain proper structure during processing and trafficking to the host cell membrane and proper function of mediating cellular receptor binding and fusion of the virion membrane with the target cell membrane.

*Variation in V2 and at the base of the V3 loop.* Outside of the V1 and V4 loops, which exhibited the highest levels of change, other sites of notable variation included the V2 loop and the base of the V3 loop. The emergence of adaptations in the V2 loop involved N-linked glycosylation sites (NLGS), in both animals Mm 10-10 and Mm 198-08 (Figure 21). Although, variation differed between Mm 10-10 and Mm 198-08, we found the evolution of changes resulting in a loss of a NLGS in the V2 loop. For Mm 10-10 we observed a decrease in N202 from weeks 18 to weeks 25 post-infection, with the concurrent emergence of a lysine, and later a serine at this position (Figure 21A; right panel). However, at 29 weeks p.i. the frequency of asparagine increased to 88.88% at position 202. Asparagine was never lost from the population, therefore, the NLGS at position 202 remained in the virus population over the course of the infection, however, a reduction at position N202 could be interpreted as a loss in the frequency of variants in the population that contained an N-linked glycan at position 202. In Mm 198-08 we found a similar pattern of change with asparagine at position 198 (N198), which contains an NLGS. Similar to N202 in Mm 1010, we found a decrease in the frequency of N198 from 100% to 64.01% beginning at week 18 to week 25, which was then followed by an increase in N198 to 84.52% by week 29 p.i. (Figure 21B; left panel). Interestingly, in Mm 198-08 we observed the emergence of a D201E substitution that became dominant by 29 weeks p.i. Again, similar to the V1 and V4 loops we observed no changes at any of these positions in Mm 156-10 or Mm 174-08 over the course of the infection that we sampled (Figure 21 C and D).
Figure 21. Evolution of sequence diversity in SIV Env V2 loop.
The frequency of codons were determined from deep sequence reads of each sample assembled to the de novo consensus sequence of the same sample using the vphaser software program (Broad Research Institute). A-D. Three sites within V2 is plotted over the first 29 weeks of in vivo viral replication of Mm 10-10 (A) and Mm 198-08 (B), over the first 15 weeks of in vivo viral replication in Mm 156-10 (C), and over the first 13 weeks of in vivo viral replication of Mm 174-08 (D), and are representative of sites with the highest codon variation in the V2 loop.
Although, we did not identify variation in the V3 loop to the level found in the V1/V2 and V4 loops, we did find a single site at the base of the V3 loop, position 345, where a W345R substitution emerged in both Mm 10-10 and Mm 198-08 (Figure 22). In Mm 10-10, we found that two different codons for arginine emerged sometime between 10 and 14 weeks p.i., suggesting strong selection for arginine at position 345. Although, tryptophan returned to become the dominant residue at 25 weeks p.i. both arginine codons (AGG and CGG) increased to higher frequency then tryptophan by 29 weeks p.i. In Mm 198-08 the emergence of arginine (coded for by AGG) continued steadily to become the dominant residue.
by 29 weeks p.i. As in Mm 10-10 Arg (CGG) emerged but not to the level of Arg (AGG). The W345R substitution was not detected in either Mm 156-10 or Mm 174-08.

Overall, we found several sites of adaptation emerge in the gp120 subunit that clustered in two regions: the V1 and V4 loops (Figure 23). Although, no structural information exists for the SIV V1/V2 domain, the V4 loop is present in a crystal structure of the gp120 subunit [177]. The structure places

D415 near the apex of the V4 loop and P421 near the middle of the V4 loop; both are completely exposed on the higher order Env trimer (Figure 24) [177]. The D415N adaptation introduces a novel N-linked glycosylation site, which was not present in the stock population of the SIV<sub>mac251</sub> swarm. Most isolates
of SIV$_{mac}251$, similarly, do not possess a Asp residue at position 415. An adaptation at A41T in SIV$_{mac}239$ results also results in an N-linked glycosylation site at position 415. Both A41T and P421Q was previously observed in env sequences cloned from an SIV$_{mac}239$-infected rhesus macaque and found to confer resistance to high-titer neutralizing plasma against parental SIV$_{mac}239$ [162].

![Figure 24. SIVmac gp120 V4 loop is target of Env-specific antibodies leading to substitutions and in-frame deletions.](image)

A crystal structure of the SIV$_{mac}$ gp120 subunit [177] showing the presence of the V4 loop (orange) on the external surface of the gp120 monomer where substitutions and in-frame deletions emerge rapidly during early SIV-infection in macaque hosts. The D415 adaptation leads to a novel N-linked glycosylation site in the V4 region and is found near the apex of the V4 loop. P421 is a hotspot of adaptation, with several substitutions emerging at this site. The presence of a bend near the Pro site suggests adaptations may lead to a loss of rigidity within the loop that is maintained by the Pro. A R424Q change was observed in two of four animals in our cohort and occurs near the bottom of the V4 loop. The Δ symbols denote the span of in-frame multiple amino acid deletions that emerged to become dominant in two animals (Mm 10-10 and Mm 198-08) by 18-23 weeks p.i. Image produced with PDB: 3FUS.

2.5.6 Effect of antibody escape changes on Env conformation using Env-specific Rh mAbs

We found substitutions at 3 sites in gp120, which were previously identified as antibody escape adaptations to high-titer neutralizing plasma against SIV$_{mac}239$ [162], emerge within the viral populations of Mm 10-10 and Mm 198-08. One site, position 138, is in the V1 loop and two sites, positions 415 and 421, are in the V4 loop. The repeated emergence of substitutions at these sites and the antibody resistance they conferred indicated changes at these sites conferred escape to antibody either directly, by disrupting
recognition within the targeted epitope or indirectly through conformational changes that acted distal to the actual substitution. To test whether these adaptations affected antibody binding, we performed ELISA based binding assays using Rh mAbs with known target specificities.

The specificity of the panel of Env-specific mAbs was previously identified by cross-competition analysis and were categorized into several competition groups that targeted the V1, V2, V3, and V4 loops [110, 297, 298]. We found that all Rh mAbs recognized soluble SIVmac239 gp140 (Figure 25). We introduced V1 A138T and V4 A417T and P421Q into gp140 and tested our panel of Rh mAbs for binding to these mutant gp140 proteins. Surprisingly, we found that V1 A138T and V4 A417T had no detectable effect on recognition by any of the Rh mAbs despite the expectation that these mutations introduced novel glycan modifications (Figure 25). However, we did find that Env P421Q had a drastic effect on recognition to mAbs that targeted the V4 loop. The epitopes for V4 loop specific mAbs from competition groups VIb and VII are thought to be conformational [110, 297]; our results suggest that P421Q disrupted binding to the target epitope. Consistent with the possibility that P421Q affects a conformational determinant of antibody recognition, we saw no detectable effect on binding of the V4-specific Rh mAbs with A417T, which lies just four residues upstream of P421Q.

Variation at P421 has been reported for env sequences recovered from SIV-infected macaques [162, 246, 247, 256] and in our study, we found substitutions emerge by week 18 post-infection at this site including P421L, P421S, and P421T. We introduced P421L and P421S into gp140 to test whether these substitutions had a similar effect on antibody binding as P421Q. Interestingly, we found that both substitutions had a drastic effect on recognition to group VII mAbs C26 and 1.10A, similar to P421Q (Figure 26). However, group VIb mAb 3.5F recognized P421S at levels observed with control mAb 3.11H (which recognizes a linear epitope within V3). mAb 3.5F had reduced binding to P421Q, but did not bind P421S. These results provide evidence that polymorphism at P421 is likely selected for based on the differential effects these substitutions have on antibody recognition in vivo. We engineered P421A into gp140 and again tested binding to V4 loop-specific mAbs C26, 1.10A, 3.5F, and V3 loop-specific mAb 3.11H. Interestingly, we found that 3.11H and all V4-loop specific mAbs bound to gp140 P421A,
Figure 25. Env P421Q substitution leads to loss of binding of V4-loop specific mAbs to SIVmac239 gp140 soluble trimers. We tested the effect of SIVmac239 adaptations on antibody recognition by a panel of Env-specific mAbs with specificity to the V1, V3, and V4 loops. Binding was measured over serial dilutions of mAb against 50 ng of soluble gp140 WT or mutant in 96-well plates by ELISA. An HRP-conjugated anti-IgG antibody was used with TMB substrate for detection of binding. Absorbance values were read at 450 nm and experiments were performed in triplicate.
Substitutions at V4 loop P421 site lead to a loss of recognition to V4 loop-specific mAbs.

50 ng of recombinant SIV\textsubscript{mac}239 gp140 WT or mutant protein were plated in 96-well plates and incubated overnight at 4°C. Binding was measured over seven serial six-fold dilutions of mAb in triplicate at a starting concentration of 10 μg/ml. However, binding was reduced with competition group VII mAbs C26 and 1.10A (Figure 26). Taken together, our results suggest that the substitution of the proline at V4 position 421 may function primarily to alter a conformational epitope, and that the nature of the replacement residue is not as critical as the removal of the proline.

Interestingly, we confirmed that none of the V4-loop or V3-loop specific mAbs neutralized SIV\textsubscript{mac}239, as reported previously, using a standard neutralization assay with C8166-SEAP reporter cells. As Env A138T, A417T, and P421Q had only been found to provide escape to polyclonal antibody populations in high-titer neutralizing plasma, we wanted to confirm our binding results with gp140 in the context of virus infection, but could not use neutralization-resistant SIV\textsubscript{mac}239. Therefore, we proposed to introduce each of these changes into the background of the neutralization-sensitive variant SIV\textsubscript{mac}316, a macrophage-tropic virus isolated from an animal infected with SIV\textsubscript{mac}239. This isolate only contains 8 amino acid differences within Env relative to SIV\textsubscript{mac}239 but is readily neutralized by SIV\textsuperscript{+} plasma and Env-specific mAbs, including C26, 1.10A, 3.5F, and 3.11H [110, 111].
Therefore, we introduced Env A138T, A417T, and P421Q into SIV\_mac316 and tested the neutralization sensitivity of these variants against our panel of V4 and V3 loop-specific Rh mAbs. Similar to our ELISA binding results, we found P421Q was the only substitution that altered the neutralization sensitivity of SIV\_mac316 (Figure 27). While it had no detectable effect on neutralization by V3-specific Rh mAb 3.11H, the P421Q substitution resulted in increased resistance of SIV\_mac316 to neutralization by all the V4 loop-specific Rh mAbs tested, thereby confirming our ELISA binding data with gp140 P421Q. Furthermore, we found, for the first time, that Env P421Q could alter neutralization sensitivity to antibodies with known specificities providing evidence that polymorphism at P421 is likely being driven by antibodies that target the V4 loop.

Figure 27. Env V4 P421Q confers neutralization resistance to V4-loop specific mAbs when tested in the context of neutralization sensitive SIV\_mac316. The level of virus neutralization of Env-specific mAbs was measured over 10 serial 2-fold dilutions of mAb against SIV\_mac316 WT or with SIV\_mac316 P421Q in triplicate. 10 ng p27 of WT and mutant virus stock was used to infect 5 x 10\(^3\) C8166-SEAP cells/well in a 96-well plate. mAb 3.11H targets a V3 linear epitope and was used as a control. Infectivity is reported as the level of SEAP activity above background from uninfected C8166-SEAP cells.
Last, we wanted to determine whether any of our antibody escape changes had an effect on neutralization by sCD4-183, a soluble recombinant version of CD4 that contains the first 183 residues of CD4, encompassing the two N-terminal domains of CD4. We found that none of the changes altered sensitivity to sCD4-183, with all Env substitution mutants being neutralized to the same level as SIV\textsubscript{mac}239 (Figure 28). This suggests that these common substitutions that arise in vivo have little or no impact on the interaction of the Env trimer with CD4.

![Figure 28](image)

**Figure 28. SIV\textsubscript{mac}239 in vivo antibody escape adaptations do not confer resistance to neutralization by sCD4.**

The level of virus neutralization of sCD4-183 was measured over 10 serial 2-fold dilutions of mAb against SIV\textsubscript{mac}239 WT and Env point mutants in triplicate. 1.5 ng p27 of SIV\textsubscript{mac}239 virus stock was used to infect 5 x 10\textsuperscript{3} C8166-SEAP cells/well in a 96-well plate. SIV\textsubscript{mac}316 was used as a control and 10 ng p27 was used to infect 5 x 10\textsuperscript{3} C8166-SEAP cells/well in a 96-well plate. Infectivity is reported as the level of SEAP activity above background from uninfected C8166-SEAP cells.

### 2.5.7 Detection of ADCC responses in SIV\textsubscript{mac}251-infected rhesus macaques

A recent report by Alpert \textit{et al.} found that SIV\textsubscript{mac} infected animals mount robust ADCC responses under both vaccination regimens or during experimental infection [284]. These results raise the possibility that ADCC activity could drive selection of adaptations in Env \textit{in vivo}. In order to ask whether the common changes we observed in V1 and V4 influence sensitivity to ADCC, we first had to establish whether any of the macaques in our cohort had detectable ADCC responses. Using a recently published ADCC assay that takes advantage of two engineered cell lines, a SIV/HIV target cell line and an NK effector cell line, providing a high resolution, standardized assessment of ADCC responses, we tested plasma from the terminal timepoint of each animal for ADCC activity against SIV\textsubscript{mac}239 [299]. Similar to antibody binding and neutralization data, we found that Mm 10-10, Mm 198-08, and Mm 174-08 had
detectable ADCC responses to SIV<sub>mac</sub>239, but Mm 156-10 did not (Figure 29). We also tested plasma from day of challenge for two animals, Mm 10-10 and Mm 198-08, we found no ADCC activity (data not shown).

2.5.8 Individual adaptations do not confer resistance to ADCC activity in SIV<sub>mac</sub>251-infected rhesus macaques

The finding of detectable ADCC responses in our cohort of animals led us to ask whether adaptations that emerged in these animals may have provided escape to ADCC. Many of these same changes, in isolation, have been shown to confer resistance to neutralization [162], and were also sufficient to confer resistance to neutralization in the context of SIVmac316 (Figure 27). Thus, we tested V1 adaptation A138T and V4 adaptations A417T and P421Q in the ADCC assay, and found that none of these changes alone were sufficient to confer resistance to the ADCC activity found in week 29 plasma (Figure 29). Similarly, we selected two V4 loop deletions to test for escape to ADCC. Again, we found that neither deletion conferred resistance to ADCC activity of week 29 plasma (Figure 31). Thus, while individual changes were sufficient to reduce sensitivity to neutralization, they didn’t result in a
significant reduction in ADCC. This may mean that ADCC activity in plasma is not directed towards the V1 or V4 epitopes, or alternatively, that the activity represents multiple independent targets such that no individual change is sufficient to confer a measurable reduction in ADCC sensitivity.

Figure 30. SIVmac239 Env neutralization escape mutants do not confer resistance to ADCC.
CEM.NKR CD4⁺ CCR5⁺ T-cells were spinoculated with 200 ng p27 of SIVmac239 or single point Env mutants. On day 4 infected target cells were incubated with NK effector cells expressing Rh CD-16. ADCC activity was measured over 8 serial two-fold dilutions of SIV⁺ plasma incubated with target cell:effector cell mixtures for 8 hours at 37°C. The level of infectivity is reported as luciferase activity above the background of uninfected target cells and normalized to infected target cell:effector cell mixtures that did not receive plasma. Mixture of target cells and effector cells in the presence or absence of plasma was done in 96-well plates and performed in triplicate.
Figure 31. SIVmac in vivo Env V4 deletions to not confer resistance to ADCC.
CEM.NKR CD4⁺ CCR5⁺ T-cells were spinoculated with 200 ng p27 of SIVmac239 or Env V4 deletion mutants. On day 4 infected target cells were incubated with NK effector cells expressing Rh CD-16. ADCC activity was measured over 8 serial two-fold dilutions of SIV⁺ plasma incubated with target cell:effector cell mixtures for 8 hours at 37°C. The level of infectivity is reported as luciferase activity above the background of uninfected target cells and normalized to infected target cell:effector cell mixtures that did not receive plasma. Mixture of target cells and effector cells in the presence or absence of plasma was done in 96-well plates and performed in triplicate.
2.6 Discussion

Primate lentiviruses establish persistent infections through several immune evasion mechanisms to mediate continuous uninterrupted replication. Evasion of antibodies has selected for extreme genetic variation and inherent Env features leading to among other things, neutralization resistance. SIV_{mac} strains are typically considered neutralization resistant strains similar to primary circulating strains of HIV-1, yet SIV env sequences isolated from macaque monkeys universally have a high degree of genetic variation often clustered in variable loop regions of gp120, including the V1/V2 and V4 loop, as well as the ectodomain of gp41, raising the questions 1) what is the evolutionary dynamics of SIV populations under detectable antibody-mediated selection over the course of infection, and 2) why do we observe such strong genetic variation in the absence of detectable neutralization? To uncover in vivo SIV env evolution during acute and early SIV infection we deep sequenced the env gene from a cohort of four SIV_{mac}251-infected rhesus macaques over the first 13 to 29 weeks of viral replication and from the uncloned SIV_{mac}251 swarm inoculum.

We focused on defining the early genetic events within the SIV env gene, which codes for the Env glycoprotein and is the primary target of SIV-specific antibodies. Our analysis was based on deep sequencing of longitudinal samples from SIV-infected rhesus macaques. This experimental approach provided two advantages to uncover the evolution of Env adaptations critical for antibody escape. First, access to the virus inoculum (a swarm stock of SIV_{mac}251) allowed us to unambiguously define the genetic diversity within the env gene of the stock population. Second, well-spaced longitudinal plasma samples provided the opportunity to accurately track env sequence evolution during early SIV infection in animals that had been inoculated in parallel, with the same dose, and route of infection. We measured we measured the Env-specific antibody response in 3 of 4 animals in our cohort. One animal (Mm 156-10) did not mount a detectable antibody response, and the absence of an Env-specific antibody response in Mm 156-10 correlated with high viral loads and rapid progression to disease. Another animal (Mm 174-08) mounted an Env-specific antibody response, but viral loads were quickly suppressed to low levels, making it difficult to track sequence evolution after week 13 p.i. Therefore, we uncovered the
earliest SIV env sequence changes starting from an uncloned SIV\textsubscript{mac}251 stock inoculum through transmission in all four animals, and in two of the four animals we tracked env evolution up through 29 post-infection.

Second, Illumina deep sequencing provided the resolution to detect minor variants within the viral population as they emerged during the course of infection in each animal and track adaptations over time providing a comprehensive analysis of sequence variation at the population level. Studies of viral escape to CTL during early infection using NGS provided strong evidence of the capacity of HIV and SIV to produce variation that allows for selection of escape from adaptive immune selection [288-291]. Another study focusing on acute and early HIV-1 evolution by deep sequence analysis found a similarly homogenous population early in the infection until the emergence of changes detectable as early as 59 days post-presentation [291]. They attributed many of the adaptations they observed throughout the genome (up to 165 days post-presentation) to CTL immune responses as many of the adaptations occurred in CTL epitopes. Thus, deep sequencing methods provide a level of resolution to detect adaptations as they emerge from the single transmitter/founder (T/F) lineage during transmission.

Here, we defined Env diversity of the SIV\textsubscript{mac}251 stock inoculum prior to transmission. We found the diversity was present in both subunits of Env and the level of diversity was similar to other uncloned stocks of SIV\textsubscript{mac}251 measured by SGA [304]. Transmission of HIV-1 undergoes a similar transmission bottleneck from the donor host to the newly infected recipient host [207, 208]. The SIV/macaque model confirmed these findings and found mucosal transmission of SIV to undergo a similar genetic bottleneck as primary HIV-1 infection and found that mucosal transmission posed a more severe barrier relative to intra-venous (i.v.) transmission [210]. Thus, it is well documented that mucosal transmission of HIV and SIV leading to infection is established by a small number, or even a single, genetic lineage, or variant [207, 208, 210], in contrast to as many as nine T/F variants detected during i.v. SIV inoculation of macaques [210].

As our animals were inoculated by low-dose mucosal (rectal) challenge, we expected a genetic bottleneck event because of the mucosal barrier that must be crossed in order to establish infection. A
recent report addressed whether transmission of the transmitter/founder variant occurs through a
stochastic process or if certain variants within the population of the donor have a selective advantage in
the environment encountered during mucosal transmission. They concluded that transmission appeared to
select for variants with a viral fitness advantage [307]. Here, we made two primary observations relevant
to SIV mucosal transmission. One, we found the T/F variant in each animal did not possess two amino
acids within the V4 loop that were present in the population of the stock inoculum. Three of the four
animals, similarly, did not have a three amino acid insertion in the V1 loop that was present in the stock
population. We do not know, with any statistical certainty, whether a shorter V1 or V4 loop provided a
selective advantage during transmission through a fitness advantage. Two studies of early HIV Env
variants found a preference for shorter V1/V2 loops [308, 309], however, a recent analysis investigating
signatures of HIV/SIV transmission in SGA sequences of several SIV and SHIV cohorts found shorter V1
and V4 loop regions in some SIV cohorts, but not all [310]. Interestingly, this recent study identified 4
signatures present in the C1 region of SIV Env that they identified in SIV$_{sm}$E660 cohorts. However,
SIV$_{mac}$251 stocks in the Gonzalez et al. study and, in our study, contained all of the signature residues
within the stock and the earliest sampled timepoints, unlike the SIV$_{sm}$E660 swarms they analyzed. As
SIV$_{sm}$E660 is not completely adapted to macaques, whether these signatures are relevant for transmission,
or for adaptation required in the macaque host, was not distinguished. Second, we observed several
residues present at frequencies of 5%, or lower, in the stock population were the dominant residue at the
earliest timepoint in Mm 10-10 and Mm 198-08, suggesting the transmission of a rare variant in these
animals. Whether, the residues present a selective fitness advantage, as suggest by Carlson et al., is
unknown, however, many residues were present in the Thr/Ser stretch of V1. Testing these residues using
our Fit-Seq assay would help to establish whether these changes have any relative fitness differences to
SIV$_{mac}$239.

Illumina deep sequencing provided the sensitivity to capture the emergence of diversity from the
initially homogenous viral populations observed 2 or 3 weeks into infection. Furthermore, the resolution
of deep sequencing data allowed us to define the emergence of adaptations relative to the entire viral
population. We observed the emergence of sequence diversity as early as six weeks p.i. in the V1 loop of two animals in the cohort, Mm 10-10 and 198-08. We have no indication whether the blips in variation occurred stochastically or were selected for in each animal, however, the similarity in Env region and time, would suggest a host selection pressure. Although, the animals in our cohort were not positive for protective MHC alleles (Mamu A*01 or Mamu B*17), three of the four animals were positive for Mamu A*08. Mamu A*08 was found to target a linear peptide mapping to positions 117-124 of the V1 loop of HIV-1. The N-terminal anchor residue for the viral peptide is a lysine. In our cohort, the V1 loop K120R change was among the sites of variation during the 6-week blip in both Mm 10-10 and 198-08 (Figure 14 and Figure 15). A similar adaptation, R120K, was recently reported as an escape change to a CD8+ T cell epitope in SIVmac239 Δnef-vaccinated rhesus macaque that expressed over 8 MHC alleles [311]. Therefore, it is possible that the earliest detection of variation at position 120 in both animals in our cohort was due to CTL targeting driving rapid selection of escape. This prediction is consistent with the model of similar changes leading to viral escape in animals with the same or similar MHC haplotypes, yet the absence of these changes in animals that do not express the same MHC alleles.

This blip in diversity was followed by several weeks of low level sequence change within the viral populations of all animals until ~10-14 weeks p.i. when we detected the emergence of adaptations in several regions of gp120 including the V1/V2 loops, the base of the V3 loop, and the V4 loop. The evolution of adaptations in the gp120 correlated with the onset of antibody responses in our animals and supported several reports of SIV Env-specific antibody responses targeting the V1 and V4 loops [162, 249, 251-255, 312, 313]. The V1 and V4 loops are primary sites of antibody escape [251, 312-314]. More recently reports testing Env-based vaccine approaches find antibody responses specific for variable loop regions, despite the lack of complete protection [314, 315].

A study from our lab identified a rhesus macaque that developed a high-titer neutralizing antibody response to the parental SIVmac239, yet still possessed a typical viral-load course and succumbed to AIDS-related symptoms. This case study provided a rare opportunity to study antibody escape in vivo and found changes that conferred complete neutralization resistance to the high-titer neutralizing plasma
mapped to specific adaptations in V1 and V4 and partial neutralization resistance mapped to the membrane proximal external region (MPER) of gp41. Studies from SIVmac251 infected macaques similarly showed that adaptations in the V4 loop, including in-frame deletions, increased neutralization resistance to autologous SIV+ plasma [255].

Interestingly, we did not observe extensive env variation in two animals in our cohort (Mm 156-10 and Mm 174-08). Mm 156-10 had no detectable antibody response and maintained high viral loads typical of a rapid progressor, eventually succumbing to disease by 15 weeks p.i. The lack of a detectable antibody response provided strong evidence that Env variation is largely driven by antibody-based immune selection and in the absence of antibody, the virus population is able to replicate unrestricted to extremely high titers. Of note, the swarm stock did not appear to possess a large proportion of variants with extremely pathogenic variants responsible for rapid progression to disease, as we did not observe the same pathogenesis in the other animals in the cohort.

In contrast, Mm 174-08 did have antibody responses similar to Mm 198-08 and Mm 10-10, yet we failed to detect any variation up to 13 weeks p.i. We failed to amplify viral RNA after 13 weeks p.i. using our two-amplicon system, but were successful in amplifying viral RNA from plasma at four timepoints (17, 22, 24, and 28 weeks p.i.) using V1 and V4 regions-specific primers that were subjected to deep sequencing. Although, our analysis is still ongoing, it appears that env variation in this animal may remained low throughout early infection. When we tested IFN-γ activity of PBMCs isolated from Mm 174-08, we detected cytotoxic T-lymphocyte (CTL) responses directed to Gag and Env by ELISPOT (data not shown). Thus, the presence of low viral loads after peak of viral replication during acute infection and the detection of both cellular and antibody-based humoral adaptive immune responses suggests Mm 174-08 was able to mount an adaptive immune response leading to control of viral replication, resulting in low Env diversity.

We found a striking level of threonine, serine, and alanine substitutions evolve in the Thr/Ser repeat region of V1 resulting in the rapid emergence of multiple amino acid insertions from 18 weeks p.i. thru to 29 weeks p.i. in Mm 198-08. Multiple amino acid insertions consisting of threonine, serine, and
alanine residues is consistent with the selection of O-linked glycosylation attachment sites that are predicted to exist in SIV\textsubscript{mac} and SIV\textsubscript{sm} strains [260]. The V1 A138T adaptation was the only site in V1 found to confer neutralization resistance in the previously mentioned study and here this adaptation emerged in both Mm 198-08 and Mm 10-10. Interestingly, in Mm 10-10 A138T did not emerge until 29 weeks p.i., in parallel with at T136N substitution, resulting in the introduction of a novel NLGS. The observation of two mutations within two different codons resulting in a novel NLGS reflects the extent of selective pressure mediated by antibodies that target V1 and the selection of adaptations contributing to the glycan shield of PLV Env glycoproteins.

The emergence of variation in V1 correlated with antibody responses and consisted of substitutions but also large in-frame insertions followed by in-frame deletions. The Thr/Ser stretch displayed such high repetitive nucleotide sequence it appeared to lead to RT slippage events, which introduced multiple amino acid insertions (it was clear from comparing longitudinal consensus sequences that insertions were the result of sequence duplications). Thus, the SIV\textsubscript{mac} V1 loop may provide an evolutionary answer to deal with antibodies consistently arising and effectively inhibiting viral replication through neutralizing and non-neutralizing effector functions. The sequence plasticity of V1 is likely a direct result of the highly repetitive ACA, TCA, GCA, and CCA stretch of codons located within the actual unstructured “loop” region of V1. The highly similar codons potentially provide a mechanism of introducing amino acid changes without drastic effects on structure or function yet confer escape. The repetitiveness of this stretch of sequence potentially serves as a template that increases the probability that RT slips off and then slides back resulting in the introduction of additional sequence.

The finding that V2 loop Env-binding nonneutralizing antibodies correlated with a 30% protection of vaccinated individuals in the RV144 trial has triggered a renewed interest in the relevance of Env-specific nonneutralizing antibodies. A vaccine strategy that mimicked the prime and boost strategy of the RV144 trial in rhesus macaques challenged with SIV\textsubscript{mac}251 found the few protected animals had high avidity antibodies to the V1/V2 loops providing further evidence for a protective role of V1/V2 loop nonneutralizing antibodies [315]. Although, much focus has been given to potent bNAbs, with good
reason, it could be that V1/V2 loop Env-binding nonneutralizing antibodies actually pose a higher selective pressure than their more rare bNAbs counterparts, as they consistently arise in SIV-infected macaques and HIV-infected patients because of the exposure of the V1/V2 domain at the apex of the Env trimer.

Similar to the V1/V2 loop, the V4 loop is highly exposed on the gp120 monomer and the Env trimer. We observed the emergence of adaptations in the V4 loop that followed a similar pattern of variation as the V1 loop. We found substitutions were largely limited to 3 sites within V4. First, is an adaptation leading to a novel NLGS at position 415. D415N introduces a novel NLG site in V4, which is significant because no other NLG site exists in the V4 loop of SIVmac viruses. Interestingly, SIVmac239 and SIVmac251 differ at positions 415 to 418 with 239 viruses containing 415NTAN418 and 251 viruses containing 415DLTT418 (refer to Figure 19). In our cohort of SIVmac251-infected animals we found the introduction of the NLGS through the D415N substitution, introducing the Asn residue of the NXS/T motif. In SIVmac239-infected animals the typical substitution observed is A417T, which introduces the Thr of the NXS/T motif. Thus, the two substitutions represent a scenario of convergent evolution resulting in a glycan modification likely critical to antibody escape. the A417T (NXS/T) adaptation was previously found to confer resistance to the high-titer neutralizing activity of SIV+ plasma [162]. Interestingly, all of the V4 loop mAbs we tested bound to recombinant SIVmac239 gp140 with the A417T mutation as WT gp140. It is possible that our recombinant proteins are not being modified by N-linked glycans, however, it is surprising that this mutation did not affect antibody binding. Additionally, no SIV V4 loop-specific mAbs that are affected by the NLG at position 415 have been reported.

Perhaps the most interesting variation occurred at the highly polymorphic site P421, where several substitutions emerged in Mm 10-10 and 198-08. Several adaptations have been observed at P421 including Pro → Ser, Leu, Gln, Thr, and Ala. One substitution, P421Q, was another adaptation found to confer resistance to the same high-titer neutralizing SIV+ plasma [162]. In animal Mm 198-08, two adaptations emerged P421L and P421S, however, both remained at low frequencies. In Mm 10-10, we saw four adaptations emerge in parallel: P421L, P421S, P421T, and P421Q. The variation at position 421
is remarkable. P421 exists near the middle of the V4 loop and likely is responsible for the bend in the loop observed in the crystal structure of the SIV gp120 subunit (Figure 24 and [177]). Thus, substitutions at P421 may result in a loss of constraint leading to a loss of antibody recognition. Adaptations at P421 were found to disrupt antibody recognition to several V4 loop-specific Rh mAbs (Figure 25 and Figure 26) providing evidence that variation at P421 affects antibody recognition. Because SIV_{mac239} is already highly resistant to neutralization, we could not test the effect of P421Q on neutralization resistance with our panel of V4 loop-specific mAbs using SIV_{mac239} itself; however, we were able to show that P421Q increased the level of neutralization resistance of SIV_{mac316}, a macrophage tropic variant that is otherwise highly neutralization sensitive.

All together, we have found that antibody selection that targets the V4 loop drives antigenic variation at P421 leading to escape from antibody. Moreover, we observed that multiple amino-acid in-frame deletions typically span or flank the P421 site. We found that V4-loop deletions became dominant in the viral population of both Mm 198-08 and Mm 10-10. V4 loop deletions were similarly observed in primary env sequences we cloned from these animals and from SIV_{mac239}-infected animals (data not shown). Although, we did not show the effect of these deletions on antibody binding, others have previously found similar V4-loop deletions lead to a loss of binding to Env-specific mAbs [249]. Altogether, it is clear that the V4 loop constitutes a major antibody target in SIV Env, as concluded in other studies, however, it remains to be investigated what type of antibodies are driving selection of deletion variants. Here, we cloned two of the deletion observed in our cohort into SIV_{mac239} and tested whether they affected sensitivity to ADCC activity from plasma of animals in our cohort. Interestingly, we found they were much more sensitive to ADCC responses relative to SIV_{mac239}. We now have the opportunity to test whether such large deletions have any effect on replicative fitness. This type of analysis will shed some light on the potential fitness costs associated with such rapid protein change.

We did not observe substantial variation in the V3 loop itself consistent with several reports of SIV Env variation [162, 246, 247], but did find variation emerge at position 345 at the base of the V3 loop in the form of a W345R substitution. The W345R substitution was found in HIV-2 Env sequences
and an alanine-substitution mutant, W345A, was found to confer neutralization resistance to several distinct groups of mAbs targeting the base of the V3 loop [316]. The evolution of similar changes among HIV-2 and SIVmac strains could be due to their shared evolutionary history as both strains emerged as cross-species transmissions of SIVsm from sooty mangabeys into humans and macaques, respectively. The study of HIV-2 sensitivity to mAbs that recognize HIV-2 Env, suggested that differences in trimer configuration were responsible for increased antibody accessibility to the V3 loop in HIV-2 relative to HIV-1 [316]. Interestingly, comparison of HIV-1 and HIV-2 Env glycoproteins found HIV-2 much more sensitive to neutralization by sCD4, HIV-2+ plasma, and HIV-2 Env-specific mAbs with defined epitope specificities relative to HIV-1 Env [277, 316]. However, SIVmac strains are inherently neutralization resistant similar to circulating strains of HIV-1, indicating identical or similar adaptations emerge in PLVs regardless of their inherent neutralization sensitivities. Furthermore, the observation of the same amino acid substitution in the same structural location is further evidence of a common pattern of adaptations shared among PLVs in different primate hosts. The observation that a W345A substitution provided potent resistance to many different groups of mAbs targeting the base of the V3 loop suggests PLVs evolve along shared sequence pathways that lead to antibody escape and viral persistence [316]. NLG sites at the base of the V3 loop in HIV-1 comprise the epitope to the PGT class of broadly neutralizing antibodies providing further evidence that this region of Env is a common target amongst PLVs [146].

The significance of the evolutionary dynamics that we observed in the V1 and V4 loops is that these animals had robust antibody responses, but not high-titer neutralizing responses. Although, plasma from these animals neutralized a lab-adapted stock of SIVmac251, it did not neutralize the SIVmac251 swarm stock (data not shown). Therefore, the antibody responses in our cohort fall under the category of Env-binding nonneutralizing antibody responses [88]. The relevance of antibody responses similar to those observed in our cohort are being held under a different light with the results of the RV144 trial, which found a 30% protection rate in individuals that received the prime/boost regime [152], and subsequently found that nonneutralizing antibodies that targeted the V2 loops correlated with protection
Surprisingly, neutralizing antibodies and CTL responses failed to correlate with protection [154]. Thus, efforts are now directed at defining what effector functions these nonneutralizing antibodies may have that are leading to partial protection, in attempts to increase their protective efficacy.

One such effector function being investigated heavily is ADCC, which occurs through the direct interaction of the Fc region of antibodies with a receptor (CD16) present on NK cells leading to NK cell mediated killing of cells bound by antibody. A recent study from the Evans group found that ADCC responses were present in animals vaccinated with live-attenuated SIV and developed to a higher titer than neutralizing responses [284]. The authors suggested that the sterilizing protection observed in these animals was, at least partially, due to ADCC responses in the absence of neutralization. This assay was similarly used to test for correlation of ADCC with protection in the RV144 trial. Although, ADCC responses did not correlate with protection it did positively trend in individuals who showed protection [154].

Here, we used the same ADCC assay and found that the animals in our cohort who had detectable antibody responses did have ADCC responses, when tested against SIV<sub>mac</sub>239. We were interested to find whether adaptations that conferred neutralization resistance also conferred resistance to ADCC responses in our animals. Surprisingly, we found that individual adaptations, including in-frame deletions cloned into the V4 loop of SIV<sub>mac</sub>239, failed to confer resistance to ADCC. These results suggest that true escape from ADCC may require the multiple adaptations typically observed in primary env strains isolated from HIV-infected individuals and SIV-infected primates, as proposed for escape from neutralization [119].

Further questions that arise are why do we observe such drastic variation in the V1 and V4 loops? Secondly, do the V1 and V4 loops constitute two separate targets on the Env trimer or do they similarly affect an antibody epitope that is distal to both loops? Outside of the CD4 binding site and the MPER of gp41, HIV-1 bNAbS target the V1/V2 loops and the V3 loops, but not the V4 loop. Furthermore, the epitopes of these bNAbS are typically glycan dependent. The Thr/Ser stretch of the V1 loop is more typical of SIV<sub>mac</sub>, SIV<sub>sm</sub>, and HIV-2 Env, but not HIV-1 Env. An analysis of predicted O-linked
glycosylation of Env found no potential sites of attachment in HIV-1 Env, but did find that SIV\textsubscript{mac}, SIV\textsubscript{sm}, and HIV-2 Envs did have predicted sites of O-linked attachment in the Thr/Ser stretch of the V1 loop [260]. Thus, there are some differences in the nature of sequence variation and antibody epitopes between SIV and HIV-1, therefore, moving forward, it will important to determine whether the observations made here are relevant to HIV-1, or potentially provide information that is more specific to the SIV/macaque model. For example, as study of host antibody maturation leading to potent neutralization together with HIV-1 \textit{env} sequence evolution was recently reported [317]. This study found that maturation of potency did precede variation within the viral population at sites within the confirmed antibody epitope. These types of analysis using the SIV/macaque model will be helpful in elucidating the antibodies (and their specificities) responsible for driving selection in the V1 and V4 loop regions.
Chapter 3
The Relative Viral Replicative Fitness of Env Adaptations Measured Using Fit-Seq
3.1 Acknowledgements

Sergio Ita conceived of experiments, carried out Fit-Seq, ELISA-binding, and neutralization assays, collected and analyzed data, and drafted text.

This work was supported by grants AI083118 (NIH) to Dr. Welkin E. Johnson.
3.2 Abstract

HIV and SIV undergo a genetic bottleneck during transmission resulting in a homogeneous population during acute and early infection. However, the rapid replication kinetics coupled to the high mutation rate of primate lentiviruses rapidly generates a population of genetically distinct variants that is acted upon by host selection forces, including host immune responses. The ability to generate large populations of genetically distinct variants allows for escape from host immune forces and, in the context of antiretroviral therapy, the evolution of drug resistance.

However, the ability to escape from host immune forces or antiviral drugs, in most cases, comes at a viral replicative fitness cost. CTL responses targeted to conserved regions, such as Gag, exact a high viral replicative fitness cost that is typically balanced through compensatory changes.

Interestingly, very few studies have addressed mutations at sites of antibody escape using competition assays, which may be due in part to the time and labor necessary to produce mutants for analysis in competition assays. Here, we present a novel approach of using Illumina deep sequencing to the measure relative fitness of SIV Env mutants in competition experiments. Using short, targeted amplicons that span the mutation(s) of interest we directly assembled deep sequence reads that contained either WT or mutant sequence directly to the sequence of our reference strain, SIV<sub>mac</sub>239. The relative frequencies of WT and mutant reads were then plotted over time. We then used the slope of the line as a direct measure of viral fitness, described here as the relative replication capacity (RRC). Therefore, this assay has the potential to screen a large number of mutants in a high-throughput fashion without the need for viral sequence modification using a high-resolution measure of relative fitness.

Interestingly, we found that adaptations in V1 and V4 that confer antibody escape in vivo did not have a replicative fitness cost relative to SIV<sub>mac</sub>239. One adaptation in the V4 loop, P421Q, which we found not only provides escape to antibody in vivo, but also disrupts binding of V4 loop-specific mAbs to SIV<sub>mac</sub>239 gp140, and emerges repeatedly in outbred SIV-infected primates. We found that testing the relative fitness of P421Q in the presence of antibody revealed a positive fitness advantage over SIV<sub>mac</sub>239, and surprisingly, direct inhibition of SIV<sub>mac</sub>239 by a nonneutralizing antibody. Taken
together, our deep-sequencing based assay revealed inhibition of a neutralization resistant SIV\textsubscript{mac}239 that was not detected using the same mAb in a standard neutralization assay with C8166-SEAP cells.
3.3 Introduction

Primate lentiviruses generate viral populations of genetically distinct variants through a high mutation rate, rapid replication (~1-2 days), and high turnover of viral particles and CD4+ T cells infected by the virus. The mutants produced by the low-fidelity RT are acted upon by host immune pressures or by antiretroviral (ARV) drug therapy leading to the rapid selection of escape variants that evade the immune response or reduce the effectiveness of ARV drugs.

SIVs tolerated extensive genetic variation within env while constrained by maintenance of numerous, essential Env functions including: expression in the secretory pathway, proper oligomerization, processing, glycosylation, trafficking to sites of viral assembly, specific incorporation into virions, recognition of the primary receptor CD4 and coreceptors (e.g. CCR5 and CXCR4), insertion of the fusion peptide, re-folding and formation of the fusion pore. Viral escape from ARV drugs is typically found to incur significant replicative fitness costs and removal of ARV treatment quickly results in reversion at sites of escape to ARV drugs. In contrast, the observation of continuous intra-host sequence diversity throughout the course of viral infection suggest that adaptations either do not incur viral fitness costs or that Env has evolved mechanisms to balance evolution of adaptations in response to continuous Env-specific antibody responses with maintenance of Env function.

Unlike the fitness costs associated with CTL escape, the fitness costs associated with antibody escape remain largely unaddressed. Primary HIV and SIV env sequences cloned into the background of a molecular strain (i.e., HIV-1 NL4-3 or SIVmac239) typically replicate at or near the level of the reference strain [124, 162]. These primary Env variants typically confer escape to antibody when tested against matched plasma samples, however, they contain several non-synonymous and synonymous changes. Thus, the contribution of each individual change to antibody escape and the associated viral replicative fitness cost remain undefined.

Here, we define viral replicative fitness as the capacity to pass on viral genetic material to progeny virions. Although, we are ultimately interested in the absolute fitness of virus mutations in vivo, we are limited in our ability to measure absolute fitness in any capacity in vitro. Absolute viral fitness is
tied to the replicative fitness capacity of the virus including all pertinent steps of the lentiviral life cycle (entry, reverse transcription of the genome, integration, viral gene transcription and expression, assembly, and budding) in addition to host environmental factors including target cell density, target cell conditions for viral replication, the state of the immune response, and the diversity of the viral population.

Relative fitness estimates of HIV or SIV viral mutations can be made by infecting cells with the wild type (WT) reference strain and a single point mutation strain (mutant) in parallel and comparing the replication of each strain by ELISA-based assays that measure the level of p27 CA in the supernatant of infected cells or measuring the level of luciferase or SEAP expression from reporter cell lines, such as TZM-bl or C8166-SEAP cells, respectively. More quantitative methods such as real-time PCR provide more accurate estimates of replicative fitness differences, but these methods typically fail to detect the subtle, but significant, fitness differences between WT and mutant strains.

In essence, we can measure the replicative capacity of viruses under ideal cellular conditions using cell lines or primary cells, such as PBMCs. Furthermore, quantitative measurements of replicative fitness can be made by determining the relative fitness of a mutation or variant measured against the replication capacity of a genetically defined reference strain. The reference strain need not be the most-fit variant, but merely, establish an arbitrary fitness threshold with which to compare mutants or variants against, to determine their individual effect on replicative fitness.

Mixing the WT and mutant virus into a single pool and then infecting cells with the pooled inoculum is currently the optimal method of measuring relative fitness. Competition experiments are typically carried out under low multiplicity of infection (MOI) conditions to reduce the probability of recombination between the two strains and to increase the probability that the two variants are replicating independently of each other under conditions where target cells are in excess. The primary advantage of competition experiment is that the two variants (WT and mutant) replicate under identical conditions, thus any differences in replicative capacity should be solely due to the differences between the two strains.
Despite the advantages of mixed infections, competition experiments introduce another experimental consideration, which is the method used to distinguish the WT and mutant strains, which are essentially isogenic except for the mutation of interest. Several approaches exist to distinguish WT and mutant strains including peak heights in chromatographs [270], heteroduplex tracking assays [271], flow cytometry-based methods using GFP/RFP-labeled viruses [272], modified reporter gene systems [273, 274], or allele-specific quantitative real-time PCR [268, 275, 276]. Each of these methods provide their own advantages, however, they all share the same limitation. Each of these approaches is highly labor intensive and essentially limited to dual infections limiting the measurement of relative fitness in a given experiment to a single mutant or variant virus.

Here, we present a novel deep sequencing-based method to measure relative viral fitness that provides the same quantitative advantages of competition assays, but at a high resolution and high throughput format that provides a method to screen multiple variants in a single competition infection using fully replicating viruses without the need for any sequence modifications. We developed this method using short target-region specific primers that overlap mutations of interest to amplify viral RNA from supernatants of mixed virus infections (competitions) producing amplicons on the order of 150 base pairs. The short amplicons forego the need for rigorous in silico assembly and allow for the direct assembly of reads to the reference genome. The frequency of reads that contain WT sequence or mutant sequence is then tallied to determine the frequency of the mutant relative to the WT. As each read essentially becomes a score for WT or mutant virus, the depth of coverage provides a high-resolution measurement of any fitness difference between the WT and mutant virus. The frequency of the mutant (or WT) can then be plotted over time. Because the frequency of the mutant is relative to the WT, the slope of the line becomes a direct measure of the relative fitness difference between WT and mutant.

We utilized our deep sequencing based fitness assay to measure the relative fitness of a small panel of Env adaptations that emerged in our cohort of animals, are observed repeatedly in primary SIV env sequences, and found to confer neutralization resistance against high-titer neutralizing plasma against parental SIV_{mac}239, thus these adaptations are in vivo antibody escape changes.
Additionally, we performed competition experiments with WT SIV
mac239 and SIVmac239 Env
P421Q, a V4 loop adaptation, in the presence and absence of a V4-loop specific mAb to ask whether the
replication in the presence of antibody altered the relative fitness of Env P421Q compared to conditions in
the absence of antibody. The results of our competition experiments with antibody detected inhibition of
SIVmac239 that was not observed with standard neutralization assays, and highlight the possibility that
standard neutralization assays may not be sensitive enough to detect subtle, but potentially critical,
neutralization and secondly, suggest that antibodies considered Env-binding nonneutralizing antibodies
may in fact be inhibiting viral replication through neutralization of virus within the context of primary
infection early in SIV infection leading to the rapid selection of antibody escape mutants.
3.4 Materials and Methods

Cell lines

The human embryonic kidney cell line 293T/17 (HEK 293T/17) was used for transient transfection of full-length viral plasmids for production of virus stocks of SIV$_{mac}$239 and protein expression plasmids for recombinant SIV$_{mac}$239 gp140 protein. HEK293T/17 cells were obtained from American Type Culture Collection (Manassas, VA) and maintained in D10 cell culture medium consisting of Dulbecco’s modified Eagles medium (DMEM) supplemented with fetal bovine serum (FBS) at 10% of the total volume. TZM-bl cells were derived from human HeLa cervical carcinoma cells and used for titration of viral stocks. TZM-bl cells were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: TZM-bl from Dr. John C. Kappes, Dr. Xiaoyun Wu and Tranzyme Inc. [292-295] and maintained in D10. Human CD4$^+$ C8166-secreted alkaline phosphotase (C8166-SEAP) cells are C8166 cells transduced with a plasmid encoding the alkaline phosphotase gene under the transcriptional control of SIV LTR sequences, as previously described [296]. C8166-SEAP cells were used for titration of virus infectivity, virus neutralization assays, and viral fitness assays. The cells were maintained in R10 consisting of RPMI medium supplemented with FBS at 10% of the total volume. Rhesus immortalized hybridoma B-lymphocyte cell lines (BLCLs) were derived by Rhesus Epstein-Bar virus (EBV) transformation of peripheral blood B cells from rhesus monkeys infected with SIV$_{mac}$ strains, as previously described [110, 297, 298]. These cells were the source of rhesus monoclonal antibodies that recognize SIV Env and were maintained in R20 supplemented with primocin.

Cloning of SIV$_{mac}$239 Env mutants using full-length SIV$_{mac}$239 viral expression plasmids

The full-length SIV$_{mac}$239 expression plasmid p239flSpX and the subclone p239 3’ half have been described previously. Single point substitution mutations were introduced by site-directed mutagenesis into the p239 3’ half plasmid in the env open reading frame and sequenced to verify the desired mutation and the absence of unintended mutations. Full-length single substitution mutant plasmids were produced by restriction digest with SphI and PmlI of p239 3’ half mutant clones and then
purification and ligation of the 1.5-kb fragment into the same sites as p239flSpX. All full-length plasmids were sequenced to verify the absence of any unintended mutations.

**Transfection and production of virus stocks**

Virus stocks were produced by transient transfection of 5.5 x 10^6 HEK293T/17 cells with 10 μg of plasmid DNA in D3 media. Four hours post-transfection media was removed and replaced with fresh D10 media. Three days posttransfection the supernatant was harvested and centrifuged at 3000 rpm for 5 minutes to remove cellular debris. Virus containing supernatant stocks were aliquoted at 1 ml or 500 μl at immediately stored at -80°C. The virus content of the supernatant stocks was determined by an SIV p27 Antigen Capture Assay Kit (Advanced Bioscience Laboratories, Rockville, MD), following the manufacturer’s protocol.

**Virus Infectivity Assay**

For viral competition experiments we wanted to carry out viral infections at low MOIs. To define the number of infected cells per milliliter (ml) of virus stock, 20 μl of virus-containing supernatant-stock was serially diluted in 180 μl of R10 media (1:10 dilutional series). 100 μl of each diluted stock was then added to 100 μl of R10 media containing 1 x 10^3 TZM-bl cells, which contain a stably integrated Tat-responsive LTR promoter driving beta-galactosidase and luciferase activity. Infections were performed in 96-well plates. 72 hours post infection cellular supernatants were removed and the number of TZM-bl cells expressing beta-galactosidase was determined using the Beta-Galactosidase Staining Kit (Clonetech Laboratories, Mountain View, CA), following the manufacturer’s protocol. Cells positive for beta-galactosidase were counted at several dilutions to determine the number of infected cells per ml of virus stock.

**Virus neutralization assays**

Neutralization assays were performed to measure the neutralizing activity of rhesus mAbs against SIV strains SIVmac239 and SIVmac316 and to measure the effect of substitutions in SIV gp120 on neutralization sensitivity against Rh mAbs in the context of SIVmac316. Neutralization assays were performed using a C8166-derived CD4^+ T cell line transduced with a alkaline phosphotase reporter gene.
under the transcriptional control of the SIV LTR promoter, as previously described [296]. Neutralization was determined as the level of inhibition of secreted alkaline phosphatase (SEAP) activity over serial four-fold dilutions of mAb. Different amounts of virus input were used in experiments based on virus infectivity of SIVmac239 and SIVmac316. Virus equivalent to 1.5 ng p27 was used for experiments with SIVmac239 WT and 10 ng p27 was used for experiments with SIVmac316 or single substitution SIVmac316 mutants.

Neutralization assays were performed in 96 well plates using 5,000 C8166-SEAP cells per well. Cells in column 1 were incubated with no virus providing a background level of SEAP activity. Cells in column 2 received virus but no mAb providing a maximal level of SEAP activity. First, an appropriate volume of R10 media was added to all wells. Next, column 3 received mAb at a starting concentration of 10 µg/µl followed by nine four-fold serial dilutions across the plate to column 12. Virus was then added to columns 2-12 in a volume of 50 µl and the plates were incubated for 1 hour at 37˚C at 5% CO2 levels. After 1 hour 5,000 SEAP cells were added to every well in a volume of 75 µl bringing the final volume in each well to 200 µl. The level of SEAP activity in the supernatant was measured after 72 hours for wells that received SIVmac239. SEAP activity was measured after 5 days for wells that received SIVmac316 or 316 Env mutant viruses. SEAP activity was measured using the Phospha-Light SEAP Detection Kit (Applied Biosystems, Foster City, CA), according to manufacturer protocol.

Recombinant soluble gp140 protein production

A codon-optimized expression vector for soluble, recombinant SIVmac239 Env gp140 was described previously [303]. gp140 contains the gp120 subunit and the N-terminal ectodomain of gp41, but is cleaved prior to the TM region. SIVmac239 gp140 proteins were found to express as stable soluble trimer proteins, similar to membrane expressed Env trimers [303]. Soluble recombinant expression-optimized SIVmac239 gp140 WT and five gp140 single point mutants were synthesized by transfection of HEK293T/17 cells with our WT and mutant SIVmac239 gp140 expression vectors. Cellular supernatants were harvested 2 days post transfection and 5 days post transfection. Supernatants were pooled and centrifuged to remove cellular debris, passed through a 0.45 µm filter, and purified using Galanthus
nivalis Lectin-Agarose beads. To confirm expression of our WT and mutant SIV\textsubscript{mac}239 gp140 proteins we separated the purified non-biotinylated proteins on a denaturing SDS-PAGE gel in reducing conditions and probed for the presence of soluble gp140 through coomassie stain and by western blot using rhesus (Rh) mAb 3.11H which recognizes a linear epitope in the V3 loop of SIV gp120 (data not shown).

**Rhesus monoclonal antibody production from Epstein Barr virus-immortalized rhesus B cell lines**

Rh mAbs specific for the gp120 subunit of SIV Env were isolated from serum-free cellular supernatants of immortalized hybridoma B-lymphocyte cell lines (BLCLs) to perform ELISA-based binding assays against WT and mutant SIV\textsubscript{mac}239 gp140 proteins. To isolate the Rh mAbs, BLCLs were passaged over several weeks to produce 0.5 to 1 liter of supernatant. The supernatants were harvested, centrifuged twice to reduce cellular debris, passed through a 0.45 µm filter, and purified with a Protein A/Protein G resin mix. The concentration of the mAbs was determined by absorbance at 280 nm using a spectrophotometer.

**SIV\textsubscript{mac}239 gp140 binding by Rh mAbs by ELISA**

Wells of a 96-well plate were coated with 50 ngs SIV\textsubscript{mac}239 WT or mutant gp140 proteins overnight at 4°C. The following morning the plates were washed and blocked with 1X PBS with 5% FBS for 2 hours at 37°C at 5% CO\textsubscript{2}. Plates were washed again before adding a serial six-fold dilution of the mAb being tested for binding at a starting concentration of 10 µg/ul. The plates were then incubated for one hour at 37°C at 5% CO\textsubscript{2}. The plates were washed again and a HRP-conjugated anti-rhesus IgG was added as a secondary antibody. The plates were incubated again at 37°C for one hour. Detection was measured with the addition of TMB substrate followed by stop solution and absorbance was read at 450 nm.

**Deep sequencing based viral fitness assay**

All SIV\textsubscript{mac}239 viruses were produced and titrated for infectivity as described above. For viral competition experiments viral stocks of SIV\textsubscript{mac}239, our WT reference strain, were pooled with one single substitution SIV\textsubscript{mac}239 Env mutant (dual infection) immediately prior to infection of C8166 target cells at
an MOI of 0.01 or lower. Infections were performed in 24 well plates with 4 x 10^5 cells in a final volume of 2 ml of R10 media. 4 hours post-infection cells were pelleted at 1500 rpms for 5 minutes and the inoculum-containing media was removed and replaced with 2 ml of fresh media. Viral supernatants were harvested every 12 hours for 72 hours and stored at -80°C. We extracted vRNA from supernatants using the High Pure Viral Nucleic Acid Kit (Roche, Indianapolis, IN), per manufacturer protocol. vRNA was eluted in 50 µl of elution buffer, aliquoted at 10 µl and immediately stored at -80°C. Prior to RT and PCR amplification, viral RNA was DNase treated using amplification grade DNase I followed by a 10-minute incubation with EDTA (Invitrogen, Carlsbad, CA), according to manufacturer protocol. Three separate target-region specific primer pairs were used to amplify ~150 base pair amplicons that spanned the mutation of interest within the C1, V1, and V4 regions of SIV Env gp120. 5 µl of DNase-treated vRNA was reverse transcribed and PCR amplified using the QIAGEN OneStep RT-PCR kit (Qiagen, Valencia, CA). Thermal cycling conditions were 45°C for 120 minutes, 95°C for 15 minutes followed by 40 cycles of 94°C for 15 seconds, 50°C for 30 seconds, and 68°C for 6 minutes, and a final extension step of 68°C for 6 minutes. In parallel, 1 µl of DNase-treated vRNA was amplified using the Premix Taq DNA polymerase (Ex Taq version 2.0) (Takara, Mountain View, CA) as a Taq control using the same cycling conditions. An RT-PCR with water instead of template was also performed with each primer pair as a water control in parallel with each RT-PCR reaction. 5 µl of the RT-PCR and Taq reactions were analyzed for each sample on a 1% 1X TAE gel to verify the presence of product in RT-PCR reactions and no product in Taq and water control reactions. Samples were then PCR purified using the QIAquick PCR Purification Kit (Qiagen), per manufacturer protocol.

RT-PCR products from competition experiments performed in parallel were uniquely barcoded by a second round PCR using Nextera index primers (Illumina, San Diego, CA), according to manufacturer’s protocol. Thermal cycling conditions were 95°C for 3 minutes followed by 8 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, and a final extension step at 72°C for 5 minutes. Samples were PCR purified following the index PCR using the QIAquick PCR purification kit per manufacture’s protocol. Each sample was normalized using the SequalPrep Normalization Plate kit
(Invitrogen, Carlsbad, CA) and ~1 ng of DNA from each sample was pooled and subjected to deep sequencing using a 300-cycle MiSeq v2 kit (Illumina, San Diego, CA) on a Illumina MiSeq instrument to produce 150 bp paired-end reads. Reads were assembled to the SIVmac239 reference genome and the frequency of reads that mapped to the WT or to the mutant were counted to determine the percent frequency of the mutant virus and WT virus. The frequency of mutant reads was plotted over time and linear regression analysis was used to determine the slope of the mutant frequency line. As the frequency of each mutant was relative to the WT the slope of the mutant line was reported as a direct measure of relative fitness (relative replication capacity; ). A F-test to determine if the slope of the mutant line was significantly different then zero was performed. The p-value of the F-test and the $R^2$ value for each line is reported.
3.5 Results

3.5.1 Viral Fitness of in vivo SIV Env adaptations

When we tested individual V1 and V4 adaptations that arise in vivo in standard infectivity assays, we noted little or no impact on viral titer (Figure 32). Similarly, Sato et al. tested a subset of these changes (Env A138T, A417T, and P421Q) and also found little or no impact on infectivity [162]. This raises the possibility that these frequently occurring adaptations incur little or no fitness trade-offs – which could explain the rapidity and frequency with which they emerge in vivo. This is in contrast to many CTL-escape mutations, and HIV drug resistance mutations, which often incur large trade-offs in viral replicative fitness. Adaptations that have little-or-no impact on viral replication are effectively neutral in the absence of antibody, and can therefore rapidly contribute to immune escape and viral persistence. However, standard infectivity assays are usually not sensitive enough to detect small (but significant) differences in replicative fitness, and often do not measure the impact of mutations on

![Figure 32. β-galactosidase activity of 239 Env WT and mutant infected TZM-bl cells.](image-url)

TZM-bl cells were infected with virus stocks of SIV<sub>mac</sub>239 WT and single point substitution mutants over a 1:10 serial dilution series. 72 hours post-infection cells were fixed and stained with a β-gal staining solution overnight. The next morning cells were washed and the number of blue-stained cells was counted to determine virus infectivity per ng p27 of virus stock.
complete viral replication cycles. In order to ask whether frequently selected adaptations we observed in vivo incur fitness costs, and to quantify the impact of these changes on viral replication, we developed a highly sensitive viral fitness assay based on the use of deep-sequencing technology (Fit-Seq) (Figure 33).

**Figure 33. Fit-Seq: Deep sequencing based viral fitness assay.**
A. The WT and mutant viruses are pooled and the mixed inoculum is used to infect cells at a low MOI (<0.01). Cellular supernatants are harvested approximately every 12 hours over 72 hours. Viral RNA is extracted from supernatants, RT-PCR amplified using target region specific primers spanning the mutation of interest. B. Amplicon products are subjected to Illumina sequencing to produce 150 base pair paired-end reads. C. Reads are assembled to the SIV\textsubscript{mac239} reference genome. The frequency of WT or mutant nucleotide is determined at the site of mutation for each read to produce relative WT and mutant frequencies. D. Frequencies of WT and mutant reads are plotted over time. Linear regression analysis is performed to determine the slope of the line of WT and mutant frequencies. The slope of each line is used as a direct measure of the relative replication capacity (fitness) and an F-test is performed to determine if the slope of the mutant line significantly differs from zero.

We sought to use a mutant that potentially had a positive fitness affect to confirm the reliability of our assay methods. The *in vivo* mutation V67M is one mutation observed at high frequency in SIV\textsubscript{mac239}-infected rhesus macaques. In an alignment of chronic *env* sequences from SIV\textsubscript{mac239}-infected macaques, the V67M mutation is present in 45 of the 59 (75%) primary *env* sequences ([162, 246]; unpublished data). Notably, SIV\textsubscript{mac251} typically has a leucine at Env position 67 and 8 of the 59 SIV\textsubscript{mac239} *env* sequences had a leucine, evidence that valine is a suboptimal residue at position 67. The mechanism driving such strong adaptation at position 67 in SIV\textsubscript{mac239} is unknown, however, reports in HIV-1 suggest that the same N-terminal region of HIV-1 Env is important to surface subunit interaction with the transmembrane subunit [318]. We hypothesized that the strong selection for the V67M substitution *in vivo* provided a replication fitness advantage over WT SIV\textsubscript{mac239} that we could detect in our assay set up and was used in our preliminary competition experiments.
C8166 cells were infected with a mixed virus inoculum consisting of SIVmac239 (WT) and Env mutant V67M (mutant). Viral supernatants were harvested roughly every 12 hours for 3 days. We used viral RNA extracted from cellular supernatants as template for RT-PCR reactions using target region specific primers to produce 150 base pair amplicons. The samples were uniquely barcoded and deep sequenced by Illumina MiSeq. Sequence reads from each time-point were aligned to the SIVmac239 reference sequence to determine the frequency of reads with WT or mutant sequence. Both WT and the Env V67M mutant were detected in the mixed virus inoculum and the frequency of reads detected in the virus inoculum did not change drastically at 24 hours post-infection (data not shown). However, beginning at 30 hours post-infection the mutant frequency began to increase, relative to the WT, reaching a frequency of 44.7% at 72 hours post-infection (Figure 34). Using linear regression analysis we determined the slope of the mutant frequency line for each mutant. As the measure of read frequency is

![Figure 34. Env V67M mutation increases replication fitness of SIVmac239.](Image)

C8166 T-lymphocyte cells were infected with a pooled mixture of SIVmac239 and Env single point mutant V67M viral stocks at a total MOI of 0.01. vRNA from culture supernatants were RT-PCR amplified and processed for Illumina sequencing. Reads were aligned to the SIVmac239 reference genome and the percentage of reads with a WT nucleotide or mutant nucleotide at the mutation site was determined. The frequency of reads with the point mutation relative to the total number of reads were plotted over time. The results are reported as the mean (+/- SD) of duplicate experiments performed in parallel. Linear regression analysis was performed to determine the linear slope of the frequency of the mutant virus over time (reported as the relative replication capacity of the mutant virus), and the R² value of each line and the p-value of a testing whether the slope of the line is significantly different than zero was determined.
relative, the slope of the line becomes a direct measure of the relative replication capacity (RRC) or fitness. The slope of the line for Env V67M was 0.3823 ± 0.09096, which was significantly different then the slope of SIV\textsubscript{mac}239 WT (p value of 0.006; Table 2).

Table 2. Relative fitness of SIV\textsubscript{mac}239 Env V67M

<table>
<thead>
<tr>
<th></th>
<th>V67M</th>
</tr>
</thead>
<tbody>
<tr>
<td>RRC</td>
<td>0.38</td>
</tr>
<tr>
<td>R²</td>
<td>0.75</td>
</tr>
<tr>
<td>p-value</td>
<td>0.006</td>
</tr>
</tbody>
</table>

From this experiment we concluded that Illumina based massively parallel sequencing (MPS) could be used to detect individual variants within a mixed virus infection culture. Additionally we determined that the V67M mutation had a positive effect on replicative fitness, despite being present at a lower frequency in the initial viral inoculum.

To start the infection at equal amounts of particulate virus for our follow-up experiment C8166 cells were infected in quadruplicate with WT and V67M with equal amounts of p27 content. Furthermore, 2 additional ratios of WT virus to the V67M mutant were tested to determine whether the positive effect of the V67M mutation remained consistent over a range of virus inputs and additionally to rule out the possibility the positive fitness affect we detected was not due to a higher amount of infectious particles of the mutant over the WT in the mixed inoculum.

In the 1:1 competition of WT to V67M mutant we found that the frequency of V67M reads surpassed the WT between 30 and 48 hours post-infection reaching a maximum frequency of 69.3% at 60 hours post-infection (Figure 35) and again found V67M had a significant positive increase on RRC (Table 3). When the level of the V67M mutant was reduced to produce a 1:3 ratio of mutant to WT we still observed a positive effect of the V67M mutation on replication relative to WT. Of note, the input of WT SIV\textsubscript{mac}239 was held constant and the amount of mutant Env V67M was diluted 3-fold and then added to the pooled virus mixture. Only when we reduced the amount of mutant by another 10 fold did the positive effect of the V67M mutation fail to increase the frequency of mutant reads by 72 hours post-infection (data not shown).
Figure 35. Env V67M mutation increases replication fitness of SIVmac239 at two different inputs. C8166 T lymphocyte cell line were infected with a pooled mixture of SIVmac239 and Env single point mutant V67M viral stocks at a total MOI of 0.01 at two different inputs. For the 1:1 mixture an equal amount of p27 ng virus stock was used to infect cells. In the 1:3 the mutant V67M was diluted 3 fold and then mixed with WT. vRNA from culture supernatants were RT-PCR amplified and processed for Illumina sequencing. Reads were aligned to the SIVmac239 reference genome and the percentage of reads with a WT nucleotide or mutant nucleotide at the mutation site was determined. The results are reported as the mean (+/- SD) of experiments performed in quadruplicate. Linear regression analysis was performed to determine the linear slope of the frequency of the mutant virus over time (reported as the relative replication capacity of the mutant virus), and the R² value of each line and the p-value of a testing whether the slope of the line is significantly different than zero was determined.

Table 3. Relative Fitness of SIVmac239 Env V67M at two ratios of WT:mutant virus inputs

<table>
<thead>
<tr>
<th></th>
<th>1:1</th>
<th>1:3</th>
</tr>
</thead>
<tbody>
<tr>
<td>RRC</td>
<td>0.81</td>
<td>0.72</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

3.5.2 Fit-Seq assay variability

The novel approach to our competition assay is the use of MPS reads to determine replication of individual variants in a mixed infection culture. Using deep sequencing as a method to measure viral replication introduces potential sources of variation including variability in read frequencies between Illumina sequence runs and variability in sample representation due to the index barcodes used to multiplex sample libraries, potentially leading to false artifacts in read data. To address variability of read frequencies from index pairs used to multiplex samples we sequenced a set of samples from one replicate
infection processed with different index pairs. For each sample, the same RT-PCR amplicon products were used as template in PCR reactions with different index pairs (Table 4).

Table 4. Summary information from two Illumina sequence runs using the same library.

<table>
<thead>
<tr>
<th>Sample coverage (% of sequence reads per sample)</th>
<th>Time-point (hours post-infection)</th>
<th>Index pair ID for 1st sequence run</th>
<th>Index pair ID for 2nd sequence run</th>
<th>WT read frequencies</th>
<th>Mut read frequencies</th>
<th>P value (Paired T test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run 1</td>
<td>Run 2</td>
<td>Index pair ID</td>
<td>DC111, N702</td>
<td>Run 1</td>
<td>Run 2</td>
<td>DC111, N702</td>
</tr>
<tr>
<td>152,380</td>
<td>126,592</td>
<td>S504, N701</td>
<td>S504, N701</td>
<td>74.0</td>
<td>74.2</td>
<td>25.8</td>
</tr>
<tr>
<td>203,610</td>
<td>86,572</td>
<td>S507, N701</td>
<td>S507, N701</td>
<td>70.1</td>
<td>69.8</td>
<td>26.7</td>
</tr>
<tr>
<td>156,360</td>
<td>39,917</td>
<td>S504, N701</td>
<td>S504, N701</td>
<td>67.6</td>
<td>68.6</td>
<td>32.1</td>
</tr>
<tr>
<td>194,912</td>
<td>123,801</td>
<td>S507, N701</td>
<td>S507, N701</td>
<td>61.3</td>
<td>59.7</td>
<td>38.5</td>
</tr>
<tr>
<td>248,176</td>
<td>85,889</td>
<td>S503, N701</td>
<td>S503, N701</td>
<td>66.2</td>
<td>56.7</td>
<td>41.2</td>
</tr>
<tr>
<td>222,604</td>
<td>99,801</td>
<td>S507, N701</td>
<td>S507, N701</td>
<td>53.5</td>
<td>51.7</td>
<td>46.2</td>
</tr>
</tbody>
</table>

The indexed samples were run on separate MiSeq runs and the read frequencies from each run were determined for both WT and mutant variants (Figure 36). We observed very small, but not significant, differences in read frequencies between the two Illumina runs. Additionally, we did not observe any measurable differences due to different index pairs. These results confirmed that using MPS is a reliable and consistent method for measuring viral replication and the high resolution of MPS

Figure 36. Plot of reads of WT and mutant Env V67M produced from two separate Illumina runs. A library produced from the same competition experiment between mutant Env V67M versus WT SIV \(_{mac}^{239}\) was barcoded using different indices for each sample in the library and then sequenced by Illumina MiSeq. The reads were assemble to the WT reference genome and reads with WT or mutant sequence were plotted over time.
provides a powerful method of detecting individual virus variants within mixed infection cultures using fully replicating virus without any genome modifications.

3.5.3 Antibody escape adaptations are not associated with a loss of viral fitness

A study of a SIV\textsubscript{mac}239-infected rhesus macaque that mounted a potent neutralizing antibody response, yet maintained typical viral loads and progressed to disease, led to the discovery of multiple antibody escape adaptations that emerged in the virus populations of that animal [162]. A138T in the V1 loop and A417T and P421Q in the V4 loop were all found to confer resistance to the high-titer neutralizing plasma of the animals when they were introduced into SIV\textsubscript{mac}239. A138T and P421Q also emerged in Mm 10-10 and Mm 198-08. Env A417T introduces an N-linked glycosylation motif at position 415. SIV\textsubscript{mac}251 differs from SIV\textsubscript{mac}239 in the V4 loop with amino acid residues 415\textsuperscript{DLTT418} present in SIV\textsubscript{mac}251 and 415\textsuperscript{NTAN415} present in SIV\textsubscript{mac}239. D415N emerged in Mm 10-10 and Mm 198-08, resulting in the introduction of an N-linked glycosylation motif at position 415. As each adaptation results in a novel N-linked glycosylation motif at position 415, and A417T was previously identified as an antibody escape change, we included A417T in our analysis. Therefore, to test whether antibody escape changes are associated with viral replicative fitness costs we tested the relative fitness of A138T, A417T, and P421Q relative to SIV\textsubscript{mac}239.

We pooled viral stocks of SIV\textsubscript{mac}239, our WT reference strain, with each SIV\textsubscript{mac}239 Env single point mutants just prior to infection of C8166 target cells. We infected C8166 cells at an MOI of 0.01 or lower and harvested supernatants every 12 hours for 72 hours. As before, we RT-PCR amplified viral RNA extracted from supernatants and subjected the amplicons to deep sequencing.

We found that for each Env point mutant there was no difference in the relative frequency of the mutant relative to WT at any time point as we previously observed for Env V67M (Figure 37). We found the slope of the line for all three antibody escape changes very near to zero and none were significantly different then the slope of the line for WT (Table 5).
Figure 37. SIV Env in vivo adaptations do not incur a replicative fitness cost.
Viral RNA extracted from mixed infection supernatants were used as template in RT-PCR using target region specific primers. Amplicon products were barcoded to produce libraries that were deep sequenced by Illumina MiSeq. Reads were directly assembled to the SIV$_{mac239}$ reference sequence. The frequency of reads with the introduced point mutation relative to the total number of reads were plotted over time. The results are reported as the mean (+/- SD) of duplicate experiments performed in parallel. Linear regression analysis was performed to determine the linear slope of the frequency of the mutant virus over time (reported as the relative replication capacity of the mutant virus), and the $R^2$ value of each line and the p-value of a testing whether the slope of the line is significantly different than zero was determined.

Table 5. Relative fitness of SIV Env adaptations

<table>
<thead>
<tr>
<th></th>
<th>A138T</th>
<th>A417T</th>
<th>P421Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>RRC</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.39</td>
<td>0.09</td>
<td>0.33</td>
</tr>
<tr>
<td>p-value</td>
<td>0.09</td>
<td>0.47</td>
<td>0.13</td>
</tr>
</tbody>
</table>

3.5.4 Env V4 loop adaptation has fitness advantage in presence of Env-specific Rh mAb

We were interested in the result that Env P421Q had no significant difference in relative fitness to SIV$_{mac239}$ as this substitution was previously found to confer resistance to a high-titer neutralizing plasma from an SIV$_{mac239}$ infected animal [162]. Several substitutions were observed at this site in Mm 198-08, Mm 10-10, and Mm 156-10 and in other SIV Env clones isolated from SIV$_{mac}$ infected animals, evidence that P421 is a highly targeted site. We previously found that Env P421Q resulted in the loss of binding to mAbs C26, 1.10A, and 3.5F, which all recognized the V4 loop, but was recognized to the same level as SIV$_{mac239}$ gp140 by 3.11H, a mAb that recognized a linear epitope in V3, by ELISA using gp140 (Figure 38). We tested these mAbs for neutralization of SIV$_{mac239}$ using C8166-SEAP cells and
Figure 38 Env V4-loop-specific mAb C26 binds to recombinant SIVmac239 gp140 but does not neutralize SIVmac239 by standard neutralization assay.

A. 50 ng of recombinant SIVmac239 gp140 were plated in 96-well plates and incubated overnight at 4°C. Binding was measured over seven serial six-fold dilutions of mAb in triplicate at a starting concentration of 10 µg/ml. B. Neutralization assays were performed as previously reported. The level of virus inhibition of Env-specific mAbs was measured over 10 serial 2-fold dilutions against SIVmac239 in duplicate. 1.5 ng p27 of SIVmac239 virus stock was used to infect 5000 C8166-SEAP cells/well in a 96-well plate. Infectivity is reported as the level of SEAP activity of infected cells over background of uninfected C8166-SEAP cells.

found that none of these mAbs could neutralize SIVmac239 (Figure 38), as previously reported [110]. As mAb C26 bound to WT SIVmac239, but not to Env P421Q, we hypothesized that if we tested the relative fitness of Env P421Q in the presence of V4 mAb C26 we would detect a positive effect on viral replicative fitness. We infected C8166-SEAP cells with a mixed pool of SIVmac239 and Env P421Q in duplicate. Additionally, we performed 3 parallel infections of C8166-SEAP cells with SIVmac239 and Env P421Q in the presence of mAb C26 at 3 concentrations: 2 µg/ml, 10 µg/ml, and 20 µg/ml. Again we amplified viral RNA by RT-PCR. Amplicon products were uniquely indexed and sequenced by Illumina MiSeq to produce 150 bp reads which were assembled to the SIVmac239 genome. In the absence of mAb C26, we found the relative fitness of Env P421Q to SIVmac239 WT, similar to previous results.

We did not observe a RRC difference of Env P421Q to SIVmac239 WT in the presence of mAb C26 at 2 µg/ml (Figure 39). In contrast, we observed a clear advantage of Env P421Q over WT in the
presence of mAb C26 at 10 μg/ml and 20 μg/ml. The RRC of Env P421Q was 0.1693 and 0.1713 in the presence of 10 and 20 μg/ml of mAb C26, both of which were statistically significant from the relative fitness of Env P421Q in the absence of mAb C26 (Table 6). Our quantification of fitness for Env P421Q

Figure 39. Env P421Q provides a replicative fitness advantage in the presence of V4-loop specific mAb C26. Dual infections of C8166 cells was performed using SIVmac239 WT and SIVmac239 Env P421Q at an equal ratio of infectious p27 particles. In parallel replication was measured in the presence of mAb C26 at 2, 10, and 20 μg/ml. Replication was measured by Illumina deep sequencing of supernatant-extracted vRNA that was RT-PCR amplified using V4 loop specific primers. Positive samples verified by gel electrophoresis were barcoded, normalized, and ~1 ng of DNA per sample was pooled and sequenced on the MiSeq platform to produce 150 base pair reads. Sequence reads were aligned to the reference SIVmac239 genome to determine the frequency of WT and mutant reads in each sample. The results are the frequency of reads containing the Env P421Q point mutation relative to total reads and reported as the mean (+/− SEM) of duplicate mixed infections performed in parallel.

<table>
<thead>
<tr>
<th>mAb C26 (μg/ml)</th>
<th>no mAb</th>
<th>mAb C26</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RRC</td>
<td>R²</td>
</tr>
<tr>
<td></td>
<td>-0.02</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.17</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>0.17</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Table 6. Relative fitness of Env P421Q in presence and absence of V4-loop specific mAb.
detect any decrease in infectivity over a dilution of mAb C26 starting at 10 \( \mu g/ml \) suggesting that
detection of viral replication in the presence of antibody through deep sequencing of mixed infections
provides a higher resolution of viral inhibition relative to standard neutralization assays.
3.6 Discussion

Unlike the fitness costs associated with CTL escape, the fitness costs associated with antibody escape remain largely unaddressed. Primary HIV and SIV \textit{env} sequences cloned into the background of a molecular strain (i.e., HIV-1 NL4-3 or SIV\textsubscript{mac}239) typically replicate at or near the level of the reference strain [124, 162]. These primary Env variants typically confer escape to antibody when tested against matched plasma samples, however, they contain several non-synonymous and synonymous changes. Thus, the contribution of each individual change to antibody escape and the associated viral replicative fitness cost remain undefined.

We formally addressed the impact of a panel of SIV Env escape adaptations on viral relative replicative capacity to determine whether characteristic antibody resistance mutations incur a fitness cost. We used SIV\textsuperscript{mac}239, which is a genetically defined molecular clone that is highly pathogenic in macaques, routinely results in high peak viral loads, and typically induces an AIDS-like disease state. Additionally, it is the prototype SIV strain used for \textit{in vitro} biochemical, genetic, molecular, and virological studies, thus our fitness measurements here, are relevant to previous and ongoing studies using SIV\textsuperscript{mac}239 and its derivatives.

We found that the V1 loop substitution A138T and V4 loop substitutions A417T and P421Q had no detectable difference in relative fitness compared to parental SIV\textsuperscript{mac}239. Although, we only tested a subset of adaptations that rose in frequency in our longitudinal data set, each of the three mutations we tested here exist in highly targeted regions of Env, are repeatedly observed in multiple SIV\textsuperscript{mac} \textit{env} sequences, and conferred complete neutralization resistance to SIV\textsuperscript{+} plasma that neutralized SIV\textsuperscript{mac}239. The lack of any replicative fitness defects are consistent with a model in which variable loop regions tolerate extensive adaptations at specific sites because they incur no replicative fitness cost, which allows for rapid selection of equally fit (with respect to replication) escape mutations. Analogous to the maintenance of conserved receptor binding regions of Env to mediate entry, changes at these variable loop sites are repeatedly selected and maintained because they confer immune evasion without an associated fitness cost. This contrasts, for example, with escape from antiviral inhibitors or escape from
many CTL epitopes in conserved domains, which often result in poorer replication and require additional compensatory changes.

The results here are similar to previous investigation of viral infectivity of these mutants using the C8166-SEAP and TZM-bl reporter cell lines, however, our experiments here were performed with multiple parallel replicates in conditions in which the WT and mutant were replicating under identical conditions, at a low MOI (MOI of ≤0.01). These conditions provided one of the highest resolutions to detect replicative fitness differences, therefore we are confident that the absence of any difference between WT and mutant are reflective of a true lack of a replicative fitness difference.

It is important to note that the extensive sequence diversity of env is largely confined to the variable loop regions of Env. One model in which variable loops mediate the balancing of fitness with antibody escape would be that the unordered structure of the variable loops reduces the need to maintain structure, offsetting any potential fitness costs, however, structural evidence and mutational analysis do not support a model in which the variable loops are completely unrestrained. A recent atomic level structure of the HIV-1 V1/V2 loop revealed a highly unexpected ordered structure comprised of four anti-parallel beta sheets that comprised a “greek key” like beta sheet structure [319]. This greek key motif contained conserved sequences held in place by inter-strand disulfide bonds and reduced the length of the “V1 loop”, found to be located between two of the sheets, down to ~10-20 residues. Therefore, mutations within the beta sheets of the V1/V2 domain would likely reduce replicative fitness, however, this has yet to be tested formally. Deletion of the HIV-1 V1/V2 domain does not inhibit Env expression or replication in PBMCs, but it does render viruses more neutralization sensitive [320, 321]. Deletion of the V1/V2 domain of SIVmac239 results in significant replicative capacity deficits in immortalized rhesus macaque T cells and selects for compensatory changes in gp120 and gp41. The ΔV1/V2 deletion variant of SIVmac239 does replicate to high titer in macaques in vivo but is quickly controlled after acute phase peak of viral replication [322]. Thus, the V1/V2 loops are necessary to evade host immune responses that would otherwise control viral replication. Furthermore, deletion of the V4 loop in HIV-1 leads to defective processing of the gp160 precursor and no detectable Env expression from cells transfected with
env expression plasmids, thus the V4 loop is likely necessary for full replicative capacity [320]. Taken together, these observations suggest there are constraints on the degree to which variable loops are free to change in response to selection by antibody.

Based on our results, we suggest an alternate model in which viral replicative fitness costs to antibody escape are balanced by selection of mutations at specific sites within variable loop regions that incur the lowest, or no replicative fitness costs. Such sites may be “poised” for rapid change that retains replicative capacity while maximally altering antigenicity. This would support the observation of the repeated emergence of the same or similar substitutions within clustered regions of variable loops, specifically V1 and V4 of SIV Env in diverse, but related strains of SIV in different primates, including natural SIV infection of sooty mangabey monkeys. A key site to this strategy is P421 in the V4 loop (Figure 40). Adaptations at P421, including P421Q, P421S, P421L, and P421A, led to either a partial or

![Figure 40. Adaptations at P421 in V4 loop provide a mechanism for antibody escape.](image)

The left panel presents the V4 loop from a crystal structure image of the SIV Env gp120 subunit. Three sites of variation are highlighted in the V4 loop, which are located near the top (D415), middle (P421), and bottom of the V4 loop and are all highly exposed on the exterior surface of gp120. The right panel depicts several adaptations observed at position 421. Interestingly, each amino acid change requires only a single nucleotide change. P421Q confers escape to antibody binding and neutralization but does not have any associated replicative fitness costs, illustrating a potential mechanism to maintain viral replication in the face of constant antibody targeting in vivo.
complete loss of binding by several V4 loop-specific Rh mAbs. P421Q was found to confer neutralization resistance to SIV+ plasma with a high neutralization titer. Here, P421Q increased the neutralization resistance of an otherwise neutralization-sensitive SIV strain, SIV\text{mac}316, to the same V4 loop-specific Rh mAbs. The finding that P421Q virus had an equal replicative capacity to WT SIV\text{mac}239, which contains a proline at position 421, illustrates a mechanism by which adaptations that do no affect viral fitness can alter antibody recognition and provide escape to neutralization.

Our finding that SIV\text{mac}239 gp140 P421Q was not recognized by several V4 loop specific mAbs (Figure 25 and Figure 26) provided an opportunity to ask if the relative fitness of the P421Q mutant would remain the same in the presence of a mAb that recognized 239 WT, but not 239 P421Q. Here, we detected a clear impact on the replication of WT SIV\text{mac}239 relative to SIV\text{mac}239 Env P421Q under mixed infection conditions in the presence of V4 loop-specific Rh mAb C26, a result we failed to detect using a standard neutralization assay with C8166-SEAP cells (Figure 39). However, in the absence of antibody P421Q and WT had an equal replicative fitness, therefore the increase we observed was not due to any replicative differences between the two viruses. Based on the binding data, we conclude that the P421Q mutant virus was not recognized by mAb C26, whereas WT SIV\text{mac}239 was. Therefore, the positive increase in relative fitness of the P421Q mutant virus could be interpreted as inhibition of SIV\text{mac}239 by mAb C26 leading to an apparent increase in the relative frequency of the mutant. This finding was significant primarily because when we tested mAb C26 against SIV\text{mac}239 using the same cell line, C8166-SEAP, using a standard neutralization protocol we found that mAb C26 had no detectable effect on the replication of SIV\text{mac}239 (this was also reported in another study [110]). Thus, Rh mAb C26 (and others including 3.5F, 1.10A, and 3.11H) were all concluded to be Env-binding nonneutralizing mAbs. Our finding provides evidence that a Env-binding mAb initially characterized as a nonneutralizing mAb can have a measureable impact on the replication of SIV\text{mac}239.

The currently accepted paradigm for antibody responses to HIV and SIV divides antibody responses into 3 general categories: early-induced non-neutralizing Env-binding antibodies, highly strain-
specific antibodies, and potent broadly neutralizing antibodies [88]. Here, we present evidence to suggest that early-induced non-neutralizing Env-binding antibodies may in fact have an inhibitory effect on Env populations. The accepted use of standard neutralization assays using reporter cell lines to measure the level of virus inhibition by plasma/sera or mAbs may lack the sensitivity to detect neutralization that is occurring in the host. One relevant significance of this observation would be in the case of live-attenuated SIV, which is one of the only immunization methods that provides completely sterilizing protection to SIVmac challenges [156-158]. Env-specific antibody response are detected in animals vaccinated by live-attenuated SIV, however, testing the polyclonal antibody plasma of these animals against the parental challenge strain has failed to detect neutralization activity. It is possible that live-attenuated vaccination does induce a protective antibody response, but this response is difficult to detect using standard neutralization assays.
CHAPTER 4

Conclusions
4.1 Understanding viral population dynamics and viral escape at the earliest stages of SIV infection

In no natural setting of HIV or SIV infection has a human or primate host been able to produce an antibody response that is effective in controlling viral replication and preventing progression to late-stage disease, regardless of the presence of potently neutralizing antibodies, in large part due to the capacity for PLV populations to generate antibody escape mutants with no apparent loss in replicative fitness. This places a requirement for a vaccine to provide protection prior to the onset of genetic variation. Studies using the SIV/macaque model provided definite and unambiguous evidence that a host antibody response is effective in controlling viral replication, requiring the adaptive changes in env that are observed universally in viral env sequences isolated from rhesus macaques and human patients to mediate escape that sustains replication and persistence. Passive-transfer studies in the macaque model provide a foundation for the rationale that antibodies can effectively prevent infection during the earliest stages of infection, likely before the infecting population can generate enough diversity to mediate escape [233, 323-328]. Defining these early adaptive events is critical to our understanding of the initial encounter of the Env-specific antibody population with the viral population present during acute and early infection that produce these adaptations.

4.2 The advantage of NGS deep sequencing methods over standard sequencing to understand the dynamics of viral escape.

The evolution of variation in the env gene of SIV isolated from SIV-infected macaques was first reported in SIVmac239-infected rhesus macaques and SIVmne008-infected pig-tailed macaques by bulk PCR and cloning [246, 247]. Comparison of env clones isolated from plasma or sera against the sequence of the parental strain provided a method to unambiguously determine sites that varied from the source inoculum into each animal. The defined period of isolation of samples further provided a level of measure of variation over time. These methods have since been applied to several other Asian primate species infected experimentally with diverse SIV strains and similarly to natural SIV infection of sooty mangabey monkeys. However, each of these studies relied on bulk PCR and cloning of a limited number of timepoints spanning either early or chronic infection. This study progressed our understanding of SIV
env evolution in two key aspects: 1) we focused our study on well-spaced longitudinal samples spanning the early stage of infection (≤6 months), which provided a reliable method to interpret changes in sequence variation over the earliest stages of infection, and 2) we used deep sequencing methods to sequence longitudinal viral populations using viral RNA isolated from plasma. The depth of coverage provided by deep sequencing, relative to bulk PCR and standard sequencing, allowed us to identify variants as they emerged at low frequency, and coupled to longitudinal sampling, track their frequency over time. Thus, we were able to track the kinetics of adaptations in regions of SIV Env known to be targeted by antibody in vivo.

Deep sequencing of SIV RNA provided a method to understand the frequency of these variants relative to the total viral population. For instance, variants that emerged at a low frequency, increased over time, and in some instances replaced the WT residue was observed in each of the variable loop regions (Figure 14, Figure 15, Figure 18, Figure 21, and Figure 22). At other sites, particularly in the V1 loop in Mm 198-08, we observed rapid replacement of the WT residue by variants from one timepoint to the next (Figure 15). It is possible that the rapid fluctuations in the populations were due to genetic drift, however, this region of V1 is highly targeted by antibody. Many of the substitutions are repeatedly observed in SIV-infected macaques and an adaptation in this region of V1, A138T, was found to confer resistance against high-titer neutralizing plasma [162]. Additionally, many of the variants were maintained over several weeks of replication during a period in which the viral populations were likely large enough to maintain variants from strong selective forces acting upon those regions of Env. The rapid emergence of variation clustering in the V1 and V4 loops and variants rising to high frequency within the population illustrated the rapid nature in which SIV populations respond to host selective pressures during early SIV infection, that include, but are not limited to, antibody targeting. Therefore, this is one of the first studies to uncover the earliest dynamics of virus sequence evolution at the population level in the SIV env gene and to define the emergence and frequency of repeatedly observed adaptations in SIVmac-infected macaques with a focus on the early stage of infection.
4.3 P421 represents a key site to mediating antibody escape without associated viral replicative fitness costs

We found that P421 exhibited some of the greatest variation at a single site in our study with several residues emerging in parallel in Mm 10-10 during early SIV infection and the same adaptations (P421L, P421S, and P421T) arising in Mm 198-08 (Figure 18). When we tested many of the adaptations observed at P421, including P421A, P421L, P421Q, and P421S, we found they abrogated binding to several V4 loop-specific Rh mAbs (Figure 25 and Figure 26). Further we found P421Q, in the context of fully replicating SIV, could increase neutralization resistance in a neutralization-sensitive background (Figure 27) supporting an earlier report of P421Q conferring neutralization resistance to SIVmac239 against a high-titer neutralization plasma that potently neutralized SIVmac239 (the parental strain) [162]. Here we established that P421Q does not alter the replicative capacity of SIVmac239, providing evidence that antibody escape may pose a lower barrier relative to escape to CTL and antiviral drug escape. As antibody responses are not haplotype specific and occur in nearly all SIV-infected primates it is likely that SIVs had to evolve mechanisms to deal with constant antibody targeting. Several features of the Env trimer reduce access to conserved regions critical to receptor binding and the fusion machinery of gp41, including oligomerization, a high number of glycans on surface exposed regions, and surface exposed variable loop regions. Variation was found to tightly cluster in the V1 and V4 loops. One explanation why could be that sequence variation in the loops provides protection between regions not protected by oligomer interactions or conserved glycans. Similarly, adaptations such as D415N in the V4 loop functioned to introduce novel N-linked glycans that likely contributed to the glycan shield. Thus, sites such as P421 may represent the final feature necessary to deal with antibodies that target external regions of Env not occluded through glycans or oligomer interactions in the native Env trimer.

4.4 Longitudinal deep sequencing of SIV populations provides the opportunity to study other regions of the SIV genome.

Several reports commented on the need for sequence data for use in mathematical-based models of viral sequence dynamics [46, 182]. Part of the work described here was in part to use the well-spaced longitudinal samples of SIV-infected rhesus macaques to answer this need for sequence information to
understand viral population dynamics under host immune responses. We designed our primer system to cover the entire env open reading frame including the C-terminal cytoplasmic tail of gp41 and sequence 3’ to env, including the entire open reading frame of nef. Our two primer sets were designed to target conserved regions of the SIVmac genome, thus our primer sets should be functional with any SIVmac-based sequences. Therefore, we now have an opportunity to expand our initial findings of early SIV env evolution using banked plasma samples that exist from studies that require the need to include control, unvaccinated, but SIVmac-challenged animals. As these animals were challenged using conditions that control for the dosage and inoculation route, parallel infection of macaque animals using genetically defined SIV strains strengthens the conclusions researchers can make on the evolutionary dynamics of sequence change observed in these animals.

Our analysis of SIV env sequence evolution included the gp41 cytoplasmic tail and the nef open reading frame. Although, we were interested in viral adaptations relevant to antibody escape, several other selective pressures are present in the host that act upon the viral population. The gp41 CT mediates critical interactions with MA necessary for incorporation into budding viral particles [2], but also mediates important interactions with host signaling factors that are critical for viral gene expression and viral replication [329]. As Nef mediates several important functions relevant to immune evasion our data presents the opportunity to investigate variation in nef. It could be argued that SIVmac Nef is already adapted to the macaque host, however, it is possible that CTL responses targeted Nef selecting for immune escape.

4.5 Implications of virus neutralization missed by standard neutralization assays but detected by Fit-Seq assay

A standard neutralization assay consists of a cell line that is permissive for HIV or SIV infection that contains a reporter gene under the control of a LTR promoter. For the HIV field the most commonly used cell line is the TZM-bl cell line, which expresses luciferase upon virus infection with HIV and SIV. CEMx174 and C8166 cell lines were transduced with an alkaline phosphotase gene under the control of an SIV LTR promoter and is permissive for both SIV and HIV infection and are also commonly used for
neutralization assays. An appropriate amount of virus necessary to detect replication is typically mixed with serial dilutions of Env-specific mAbs or SIV$^+$ plasma for one hour to allow for antibody binding to cell-free virus. After the one-hour incubation period cells are then added to the mixture and viral replication is measured from the level of expression of the reporter (typically after 48-72 hours post-infection). Despite the sensitivity of the reporter genes, the amount of virus used in a standard neutralization assay is in the range of an MOI of 1 or higher. The classification of primary strains of SIV and HIV into tiers of neutralization sensitivity and the potency of mAbs has been determined using standard neutralization assays, therefore they are the benchmark assay for the field.

Here, we make the argument that our low MOI conditions coupled with deep sequencing as a measure of relative fitness captured inhibition of SIV$_{mac}239$ that was not detected using a standard neutralization assay. The significance of this finding is further emphasized by the fact that SIV$_{mac}239$ is held by the field to be a neutralization-resistant, highly pathogenic variant. Potent bNAbs against HIV-1 do not neutralize SIV$_{mac}239$ [259], and no mAb has been isolated from an SIV$_{mac}$-infected animal that can neutralize SIV$_{mac}239$. However, in all these conditions neutralization was assessed under standard neutralizing conditions.

So why would neutralization be missed using a standard neutralization assay, yet be detectable under our conditions? The first consideration would be the level of virus used in both conditions, or rather the MOI. In the standard assay, the number of cells is relatively low (10,000 TZM-bl cells/well or 5,000 C8166-SEAP cell/well) yet the virus inoculum is at an MOI of 1 or higher. In our Fit-Seq assay we maintained an MOI of 0.01, or lower, to inhibit the probability of recombination or the mutant and WT strain and increase the probability of independent replication of the two strains in the same well. In the host, antibody may encounter virus in compartments, which effectively concentrate antibody, leading to effective neutralization. This scenario would be similar to our conditions, but in a standard assay the amount of virus may be above a biologically relevant level.

Second, the resolution of deep sequencing of vRNA templates from cell-free virion versus the level of reporter gene activation driven by Tat from LTR sequences. One reason for a high MOI virus
inoculum is to ensure reporter expression, at a minimum, two-fold above background level expression. In contrast, deep sequencing is extremely sensitive. We found that the lower level of sensitivity for detection of replication by deep sequencing occurred at an inoculum of ~0.75 ng p27 of virus stock. Additionally, our level of coverage in most of our experiments was in the range of 50,000 to 100,000 fold coverage. This level of resolution may have provided the sensitivity to detect subtle, but significant, changes in replication. However, these subtle differences in neutralization in vivo could translate into significant inhibition. This assumption would be in line with a scenario in which adaptations would be selected for to escape antibodies that could bind and inhibit virus replication at a level just enough to be selective in the population.

These assumptions are further supported by the renewed interest in Env-binding nonneutralizing antibodies due to the results of the RV144 trial, which identified a correlation with nonneutralizing antibodies, but not neutralizing antibodies, with partial protection [330]. As these antibodies have been determined to be nonneutralizing, a search is on to determine the effector functions induced by the prime/boost regimen. Nonneutralizing antibodies are generally thought to arise early in infection [88]. One possibility is that these early nonneutralizing antibodies may be inhibiting viral replication through alternative effector functions, such as ADCC, however, our results suggest that, in fact, these antibodies may be neutralizing virus as well. Further research would be needed to confirm our results starting with the identification of other mutations that reduce antibody binding to Env-specific mAbs.

4.6 Final Remarks

In conclusion, antibody responses specific to the Env glycoprotein of SIV, HIV, and other primate lentiviruses exert a potent, inhibitory, and selective pressure on viral populations in virus-infected hosts. The long co-evolutionary history between PLVs and their primate hosts has selected several features on the Env gp120 monomer and Env trimer that function in immune evasion from constant antibody responses, including a two-receptor system, glycosylation and variable loops that shield conservative receptor binding regions in gp120. Despite these features antibodies still evolve in the host that can penetrate these defenses requiring the need for sequence variation that evades antibody through
escape at sites of antibody epitopes. However, escape, like any other sequence mutation, comes with an associated fitness difference. Here we helped to uncover some of the earliest events in env sequence variation using the SIV/macaque model, finding that changes cluster in variable loops 1 and 4, evolve in a fashion indicative of natural selection, and found a subset of these adaptations had no associated replicative fitness cost. Understanding that antibody escape must evolve from a homogenous population that is present during acute infection due to the genetic bottleneck during transmission highlights the need to understand early sequence dynamics. It is becoming well accepted that Env-binding nonneutralizing antibodies are protective, though not completely, in HIV-infected individuals and SIV-infected macaques. As these nonneutralizing antibodies potentially possess alternate effector functions such as ADCC it is similarly important to test Env adaptations for escape from antibodies with alternative effector functions. Here, we found that individual Env adaptations do not confer escape to ADCC, suggesting that escape in vivo likely requires multiple adaptations in a single variant to confer complete escape. Thus, the use of SGA, as we applied deep sequencing, on longitudinal samples using the SIV/macaque model might further inform the connection among adaptations such as those in V1 and V4 that potentially function together to provide escape. Understanding which changes are linked will help to inform therapeutic efforts to design Env-based immunogens, which are at the core of a safe and effective vaccine against HIV-1.
References


111. Means, R.E., et al., *Ability of the V3 loop of simian immunodeficiency virus to serve as a target for antibody-mediated neutralization: correlation of neutralization sensitivity,


Appendix

Permissions
Permission to use figure adapted in Figure 3.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3774301063063</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Dec 22, 2015</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Nature Publishing Group</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Nature Immunology</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Antibody responses to envelope glycoproteins in HIV-1 infection</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Dennis R Burton, John R Mascola</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>May 19, 2015</td>
</tr>
<tr>
<td>Volume number</td>
<td>16</td>
</tr>
<tr>
<td>Issue number</td>
<td>6</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a dissertation / thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/educational</td>
</tr>
<tr>
<td>Format</td>
<td>print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/Illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/Illustrations</td>
<td>1</td>
</tr>
<tr>
<td>High-res required</td>
<td>no</td>
</tr>
<tr>
<td>Figures</td>
<td>Figure 2</td>
</tr>
<tr>
<td>Author of this NPG article</td>
<td>no</td>
</tr>
<tr>
<td>Your reference number</td>
<td>None</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Adaptive Evolution of the SIV Envelope Protein During Early SIV Infection</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Jan 2016</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>140</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>
Permission to use figure adapted in Figure 4.

Permission is granted for the use of Fig 1, p.2 in the article detailed below in your PhD thesis only. Please cite this article as Cold Spring Harb Perspect Med 2012;2:a006866, with copyright to Cold Spring Harbor Laboratory Press.

Good luck with your thesis.

Carol C. Brown
Books Development, Marketing and Sales
Cold Spring Harbor Laboratory Press
500 Sunnyside Blvd
Woodbury, New York 11797
516 422 4038 ph.
516 422 4095 fx.
brown@cshl.edu
Permission to use figure adapted in Figure 5.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3775381346016</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Dec 24, 2015</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Nature Publishing Group</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Nature Reviews Genetics</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>The causes and consequences of HIV evolution</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Andrew Rambaut, David Posada, Keith A. Crandall, Edward C. Holmes</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Jan 1, 2004</td>
</tr>
<tr>
<td>Volume number</td>
<td>5</td>
</tr>
<tr>
<td>Issue number</td>
<td>1</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a dissertation / thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/educational</td>
</tr>
<tr>
<td>Format</td>
<td>print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>High-res required</td>
<td>no</td>
</tr>
<tr>
<td>Figures</td>
<td>Figure 1</td>
</tr>
<tr>
<td>Author of this NPG article</td>
<td>no</td>
</tr>
<tr>
<td>Your reference number</td>
<td>None</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Adaptive Evolution of the SIV Envelope Protein During Early SIV Infection</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Jan 2016</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>140</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>