The Role of HIV-1 Recombination and APOBEC3F/G-Mediated Hypermutation in HIV-1 Pathogenesis

by

Allison M. Land

A thesis submitted to the Faculty of Graduate Studies of
the University of Manitoba
in partial fulfillment of the degree of

Doctor of Philosophy

Department of Medical Microbiology
University of Manitoba
Winnipeg

The Role of HIV-1 Recombination and APOBEC3F/G-Mediated Hypermutation in HIV-1 Pathogenesis

By

Allison M. Land

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

Of

Doctor of Philosophy

Abstract

HIV-1, causative agent of the devastating pandemic of AIDS, shows incredible sequence variation, providing a substantial challenge for vaccine design. The aim of this thesis is to characterize HIV-1 diversity by examining levels of inter-subtype recombination and APOBEC-mediated proviral hypermutation, and determine its importance, by associating diversity with pathogenesis.

The results of this thesis were obtained by sequencing samples collected from subjects enrolled in two cohorts located in Nairobi, Kenya, composed of individuals involved in commercial sex work (CSW) and those who are not. We found, in agreement with previous studies, that CSW were more likely to be infected with recombinant virus than non-sex workers. This suggests that all individuals at high risk for HIV acquisition may be important drivers of viral diversity in the global pandemic and are thus an important target for prevention and intervention strategies.

We identified a subset of individuals with high levels of proviral HIV-1 APOBEC-mediated hypermutation, which correlated with $CD4^+$ count. This indicated that APOBEC3F/3G hypermutation may be important in controlling disease progression. Sequencing the APOBEC3G gene revealed three polymorphisms that were significantly associated with hypermutation; two of these were located in the region 5' of the APOBEC3G gene and may control expression. This data suggests that contrary to previous studies, APOBEC-mediated hypermutation may be controlled by increased activity of host APOBEC3G rather than by defects in the viral Vif. We thus suggest that increases in APOBEC3F/G activity may play a protective role in disease progression.

The exploitation of these findings may aid in the development of HIV prevention and therapeutic strategies.

Acknowledgments

There are numerous people who have helped me with this thesis; I attempt to acknowledge them here, but am conscious that this thank you inadequately matches the contribution.

First thanks to my supervisor, Dr. Frank Plummer, who has the amazing ability to fit things into the larger picture, and thus transform the feeling that one's work is trivial and unimportant into a feeling that it may actually be worthwhile.

Next thanks to my mentor, Dr. Blake Ball. He has helped me with every stage of this journey, providing honest input and allowing me to be honest in return. He has provided advice when asked for and motivation when needed; both were necessary.

Thank you to Dr. Ma Luo, my lab supervisor. She has been encouraging and supportive, from the very beginning.

Thank you to my advisory committee, especially Dr. Deb Court who has been influencing my scientific career since I was a second year undergrad, and Dr. Keith Fowke who challenges me.

Thank you to the grad students, post-docs, technicians and undergrads who have worked in the NML laboratory over the years. I loved coming to work every morning, and that has everything to do with you.

Thank you to all the grad students in the Plummer/Fowke labs. Thank you for beers and coffee. Thank you for sharing the good lab days, and especially for commiserating with me on the bad lab days. Thank you to the ones who graduated first for setting such a good example. Best of luck to the ones who will graduate after me.

Thank you to all the technicians and administrative staff at the NML and in Med Micro. Thank you for being able to help me get things done.

Finally, thank you to all the women who make this work possible. *Assante sana.*

Dedication

I would like to dedicate this thesis first to my parents, who instilled a sense of curiosity and a love of all things science in me, from an early age: you support and encourage me, put my happiness first and do all you can to help. Thank you.

Next to my husband, who has agreed to be taken along on this ride: Thank you for putting up with this life choice, which drives me crazy and makes me happy in often rapid succession. You support me in so many ways, from listening to me practice presentations countless times, to reassuring me that this is all possible. You are my rock.

Finally, to my daughter: your arrival has provided a keen motivation to wrap this up. Thank you for taking long naps when I needed to focus on writing, and for being wakeful when I needed something to pull me away. You remind me what is truly important and keep me grounded.

Table of Contents

List of Tables

List of Figures

Introduction

1.1 Origin of HIV

HIV-1 (human immunodeficiency virus 1) is a member of the *Retroviridae* family and the *Lentivirinae* genus, which is characterized by a long incubation period. It is part of the primate lentivirus group, which includes HIV-1, HIV-2 and SIV (simian immunodeficiency virus). HIV causes AIDS (acquired immunodeficiency syndrome), which is a disease of the human immune system characterized by a decrease in $CD4^+$ T cell count to less than 200 cells/μL (40). Over forty different species of African nonhuman primates are estimated to be infected with different lentiviral SIV infections (295). Unlike HIV, SIVs do not cause a significant depletion of $CD4^+$ T cells in the peripheral blood nor cause AIDS-like illness in their natural hosts (108,306). This is despite the natural SIV hosts maintaining high viral loads and a short *in vivo* lifespan of SIV-infected cells, suggesting high cellular pathogenicity (103,242,265).

Interestingly, disease does occur in non-natural hosts, such as infection with SIV of Asian macaques. Cross-species transmission of SIV has been shown to occur when primates in captivity are housed in shared quarters. SIV-harbouring sooty mangabeys infected macaques on numerous occasions prior to the identification of SIV and thus before testing and precautions would have been undertaken (63,170,180,208,274). Cross-species transmission has also doubtlessly occurred in nature, as at least eight SIVs are recombinant (i.e. derived from multiple, genetically distinct ancestors), indicating that a single host was infected with multiple viruses (7). For example, SIVcpz, which infects

chimpanzees, is a recombinant virus derived from ancestral SIVs that currently infect red-capped mangabeys and *Cercopithecus* monkeys in west-central Africa (16).

There are four subspecies of chimpanzee: *Pan troglodytes verus, P .t. vellerosus, P. t. troglodytes* and *P. t. schweinfurthii*, the latter two of which can be infected with SIVcpz (261). The closest relative of HIV-1 is SIVcpz*Ptt*, which infects the chimpanzee sub-species *Pan troglodytes troglodytes* (137). Transmission of the virus from chimpanzees to humans likely occurred parenterally when chimpanzees were butchered for bushmeat (260). Indeed, parenteral transmission from bites and other wounds may be the major route of SIV transmission in non-human primates (116). This cross-species transmission of an SIV to humans has happened at least three separate times, giving rise to three genetically distinct groups, HIV-1 groups M, N and O. HIV-1 group M and N viruses are derived from SIVcpz*Ptt*, transmitted by two geographically distinct groups of chimpanzees (137). The origin of group O HIV-1 is less well understood; the closest extant SIV viruses are found in gorillas (292). These viruses form a phylogenetic cluster intermingled with SIVcpz strains, however, suggesting that chimpanzees may have transmitted the virus to gorillas, which in turn passed the virus on to humans; alternatively, the common ancestor virus infected humans and gorillas separately (292).

HIV-1 group M is the most prevalent and widespread HIV that currently circulates in the human population. It is found worldwide and causes more than 95% of HIV infections (260). Group M is genetically diverse and can be further divided in clades or subtypes, based on significant phylogenetic clustering across the genome (Figure 1) (243). The

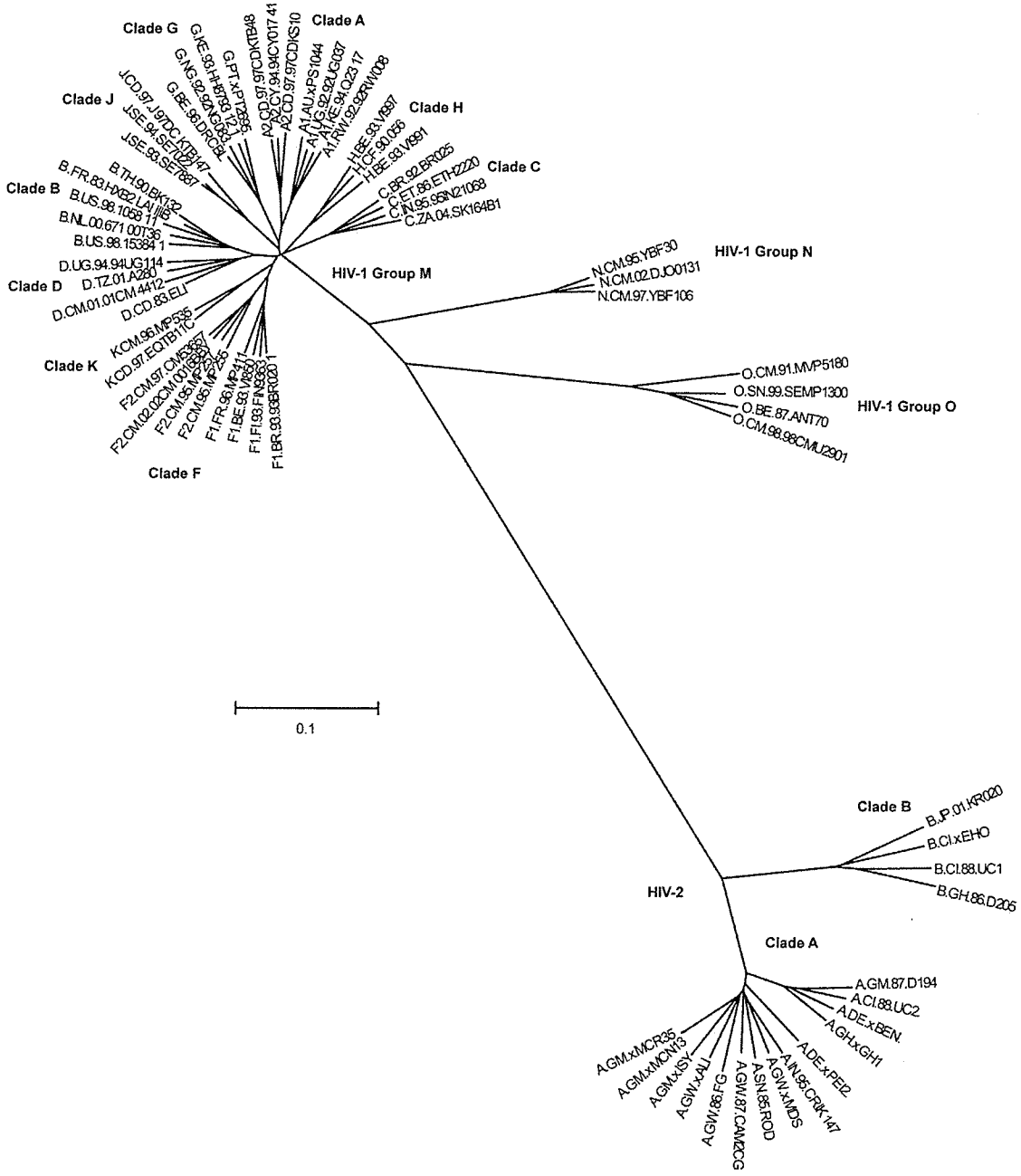

Figure 1. Neighbour-joining tree of full-length HIV-1 and HIV-2 reference sequences. Current reference sequences for HIV-1 group M, N and O and HIV-2 were obtained from the HIV Sequence Database (www.hiv.lanl.gov). The neighbour-joining tree was generated in MEGA 4 from a ClustalW alignment. Genetic distance is indicated.

clades currently identified as circulating in the global population are A, B, C, D, F, G, H, J and K. The different clades, many of which were in existence fifty years ago, likely arose due to population founder effects (260). These different viral subtypes, just like the different SIVs, can genetically recombine to form new recombinant viruses. Many different recombinant viruses have been described – over forty are actively circulating (each are found in at least three epidemiologically unlinked individuals) and are thus termed circulating recombinant forms (CRF), while many more have been described in only a single person (unique recombinant forms – URF) (189). HIV-1 group N has very limited spread, and thus far has only been isolated from individuals in Cameroon (189). HIV-1 group O is slightly more prevalent than HIV-1 group N, but is largely contained to West Central Africa (189).

The passage of SIV from a sooty mangabey (SIVsm) to humans also likely occurred by parenteral transmission (253). This cross-species transmission gave rise to HIV-2. Transmission of SIVsm to humans has occurred at least eight times, resulting in HIV-2 groups A-H, although only groups A and B show evidence of establishing epidemics (48,61,93,94,318). The distribution and prevalence of HIV-2 is limited compared to HIV-1, with the majority of infections occurring in West Africa (189).

The two oldest archived strains of HIV-1 were isolated from samples collected in 1959 and 1960 (ZR59 and DRC60, respectively). Yet, there is large genetic divergence between the two – ZR59 appears to be a subtype D virus, while DRC60 is a subtype A virus. This suggests that the HIV-1 genetic clades were already well established,

indicating that the introduction of HIV-1 into the human population is likely to have occurred at a much earlier time point. Worobey *et al.* used BEAST (Bayesian evolutionary analysis sampling trees) to determine the date of most recent common ancestor (TMRCA) for HIV-1, and found the range to be 1873-1933 (312).

For an AIDS epidemic to become established, HIV-1 likely needed a large population center to facilitate spread. Large cities were not present in central Africa before 1900, possibly explaining why the spread did not start much sooner. The earliest known strains of HIV-1 were isolated from Léopoldville (now Kinshasa, Democratic Republic of Congo) (312,337). Léopoldville was the largest city in central Africa at the time, and was on a main transportation route from Cameroon, where the SIV-infected chimpanzees were likely butchered for bushmeat and the zoonotic transmission likely occurred (260).

A new immunodeficiency was described in four American men in 1981; these were the first recognized cases of AIDS (104). A few years later, the causative agent, HIV, was isolated and identified (19,236). Subsequent studies have revealed that this viral infection was present in the population long before it was discovered. HIV was likely circulating in the United States twelve years before its discovery and in African countries clearly much longer (98).

1.2 HIV pathogenesis

HIV-1 can be spread between hosts by exposure to body fluids; significantly, via sexual intercourse (genital secretions), shared injection drug equipment (blood) and mother to

child transmission (MTCT) (blood, genital secretions and breast milk). The initial stage of infection, termed the acute stage, lasts about two months and is characterized by high levels of virus, leading to increased infectiousness (231,232,270,331). Also during this time, the $CD4^+$ cell levels drop as the virus replicates rapidly and disseminates to the various lymphoid tissue compartments (161,196). After the acute phase of HIV infection, most individuals enter an asymptomatic period where viral levels in the blood drop and reach a set point, usually below $4.3\log_{10}$ RNA copies/mL (160). This chronic phase of infection usually lasts 3 – 10 years. During this phase, HIV replication generally occurs at low levels in the lymph nodes and other tissues, and is seemingly controlled by antiviral immune responses. However, the $CD4^+$ levels continue to slowly drop and individuals not on therapy will begin to develop signs of infection and a loss of immune functions (34). In the absence of treatment at this stage, individuals can develop opportunistic infections and HIV-related cancers such as *Pneumocystis* pneumonia and Kaposi's sarcoma. $CD4^+$ levels below 200 cells/μL and these infections, known as AIDS-defining illnesses, signal the onset of AIDS (40). Destruction of $CD4^+$ T cells are a major reason for the immune dysfunction that causes AIDS. $CD4^+$ T cells are destroyed during HIV infection by both direct mechanisms, such as by the cytopathic effects of HIV, and indirect mechanisms, such as induction of apoptosis due to immune activation (4) and destruction of thymic lymphoid tissue leading to reduced production of new cells (188).

Not all individuals are equally susceptible to HIV disease. Some people, termed elite controllers, are able to suppress viral replication to non-detectable levels, in the absence

of anti-viral treatments (195). Other individuals progress slowly to AIDS and so are called slow progressors or long-term survivors (110). Still other people progress very quickly to AIDS after infection with HIV; these are referred to as rapid progressors (138). A very special subset of individuals do not become infected with HIV at all, despite high levels of exposure; these people are referred to as high-risk exposed HIV-seronegative individuals (234). There are a number of genetic factors that have been associated with differences in HIV susceptibility and disease progression. Certain HLA (human leukocyte antigen) alleles are associated with differences in disease progression as polymorphisms can affect the strength of the cellular immune anti-HIV response. Polymorphisms that affect the expression of the HIV co-receptor, CCR5, as well as certain cytokines, such as IL-10 and RANTES (regulated upon activation, normal T-cell expressed and secreted), are also associated with altered disease progression. Intracellular host antiviral factors, such as APOBEC3G, may also be important for controlling disease progression; this will discussed in detail in a subsequent section.

There is currently no HIV vaccine available, but drugs, termed antiretrovirals (ARVs) are used to inhibit HIV-1 replication and thus keep the infection under control. Different classes of drugs target various phases of the HIV replication cycle. Nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs) inhibit reverse transcription, protease inhibitors (PIs) inhibit viral assembly and entry inhibitors inhibit HIV-1 entry to the host cell. These drugs, however, are not perfect – toxicity to the host and development of viral resistance to the treatment are two

major challenges of ARV treatment, as is the fact that an estimated 6.7 million people are in need of these drugs, but are not receiving them (57).

1.3 HIV/AIDS in Sub-Saharan Africa

HIV/AIDS affects people across the globe; currently, 33 million people are infected worldwide (290). However, some areas are more affected than others – nowhere is harder hit by this epidemic than Sub-Saharan Africa, where 22 million people are currently infected with HIV (290). This region houses 67% of all people living with HIV and 75% of AIDS deaths, yet only contains just over 10% of the world's population (290). There are also other ways this region is affected: over 12 million children have been orphaned by the HIV/AIDS epidemic (290). As well, nearly 90% of children younger than fifteen living with HIV, live in sub-Saharan Africa (290). Life expectancy in sub-Saharan Africa has been dramatically affected by the HIV/AIDS epidemic; in countries with high HIV prevalence, life expectancy at birth has fallen to levels not seen since the 1950s. Indeed, the life expectancy is currently below 50 years for the region (290).

Half of all individuals living with HIV worldwide are women; however, in sub-Saharan Africa, nearly 60% of all people living with HIV are women (290). There are societal/cultural reasons for this disparity, such gender inequality and a lack of empowerment for women and girls (290). Additionally, women are actually physiologically more susceptible to the acquisition of HIV. The risk of HIV transmission is 2 – 3 times higher for a male to a female partner, than a female to a male partner

(66,216). There are a number of possible reasons for this difference. Anatomically, the female genital tract has a larger surface area and a more receptive contact surface than the male genital organs (235). Additionally, the pH of semen creates a favourable environment for HIV and prolongs viral survival (299). Also semen will remain in the female genital tract until it is absorbed, thus increasing the time of exposure. In contrast, the normally low vaginal pH makes an unfavourable environment for HIV and the duration of male exposure is determined by the duration of sexual intercourse (216). Hormonal changes, such as those caused by the use of oral contraceptive, also put women at increased risk of HIV acquisition, as does increased susceptibility to other STIs (sexually transmitted infections) (155,251).

1.4 HIV-1 Structure and Replication

The HIV-1 genome consists of two strands of positive-sense RNA. Each RNA molecule is 9.2 kb long, though the HIV-1 provirus is 9.7 kb, as during reverse transcription the ends of the RNA genome are duplicated to make LTRs (long terminal repeats) on each end. The genome contains three genes shared with all retroviruses: *gag, pol* and *env*, and an additional six genes: *nef, tat, rev, vif, vpr and vpu* (Figure 2). The RNA genome forms a complex with Nucleocapsid proteins (Gag p7), and is surrounded by a shell composed of Capsid protein (Gag p24), in a mature HIV-1 virion. Reverse transcriptase and Integrase are also associated with the ribonucleoprotein complex, as may be Protease. Matrix protein (Gag p17) forms a shell outside the core and a lipid membrane surrounds the Matrix shell. Envelope protein oligomers are associated with this membrane and dominate the surface of the HIV-1 virion (Figure 2) (56).