

ABSTRACT

Title of Dissertation: *IN VITRO* SYNTHESIS OF LONG
REVERSE TRANSCRIPTION PRODUCTS
FROM GENOMIC RNA OF HUMAN
IMMUNODEFICIENCY VIRUS

Reshma Merin Anthony, Doctor of Philosophy, 2006

Dissertation directed by: Dr. Jeffrey J. DeStefano, Associate Professor
Department of Cell Biology and Molecular Genetics

The retroviral reverse transcription reaction normally occurs in capsid-like structures in the cytoplasm of infected cells. Reverse transcription can also be carried out *in vitro* in totally reconstituted reactions with purified enzymes and model RNA templates. However, in this case fully synthesized DNAs are rarely generated from genomic RNA. This could be because the capsid creates an extremely concentrated and specific environment that cannot be completely reproduced *in vitro*. An *in vitro* system that closely mimics replication and that can be easily manipulated would enhance our understanding of the replication process. In this thesis report, *in vitro* reaction conditions that allowed efficient synthesis of DNA products up to 4 kb from genomic RNA segments of Human Immunodeficiency Virus (HIV) were generated. The reactions

required high amounts of HIV reverse transcriptase enzyme (RT) and nucleocapsid protein (NC) sufficient to completely coat the RNA template in the reaction. Synthesis of long DNA products required the formation of high molecular weight aggregates with nucleic acids, RT and NC. Removal of the dimerization region did not affect synthesis of long DNA products *in vitro*. Processivity of RT does not play a role in the synthesis of long DNA products. NC finger mutants lacking either finger or with the finger positions switched were all effective in synthesizing long DNA products suggesting that the aggregation/condensation activity but not the unwinding activity of NC is required for the synthesis of long DNAs *in vitro*. These results taken together, we propose that high molecular weight aggregates promote synthesis of long reverse transcription products *in vitro* by concentrating nucleic acids, RT enzyme and NC into a smaller area, thereby mimicking the role of the capsid environment within the host cell.

In addition, strand transfer assays indicate that strand transfer is the molecular mechanism involved in the synthesis of long DNAs and the rate of transfer (cross-over events per nucleotide synthesized) is higher than that found in tissue culture-based recombination assays. An *in vitro* system that closely mimics what occurs in the cell could be used to screen inhibitors on RT, NC and recombination.

IN VITRO SYNTHESIS OF LONG REVERSE TRANSCRIPTION
PRODUCTS FROM GENOMIC RNA OF HUMAN IMMUNODEFICIENCY VIRUS

by

Reshma Merin Anthony

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Dr. Jeffrey J. DeStefano, Chairman/Advisor
Dr. William J. Higgins, Dean's Representative
Dr. James N. Culver
Dr. Siba K. Samal
Dr. Vikram N. Vakharia

©Copyright by

Reshma Merin Anthony

2006

DEDICATION

This thesis is dedicated to my parents and to my brothers, Nikhil and Renil
for all their love and support

ACKNOWLEDGEMENTS

I express my deepest gratitude to my advisor Dr. Jeffrey DeStefano for his training, guidance and support during my graduate years. It was his constant encouragement and his belief in my ability that helped me go through some of the really difficult times during my project. Over the years, I have picked up two great qualities of a scientist from him – never-ending patience and rock-solid belief in one's work. Another great quality I have acquired from him that deserves mention is maintaining a balance between work and work-out. Today, I can run six miles even after a hard day's work.

My sincere thanks to my advisory committee members, Drs. William Higgins, James Culver, Vikram Vakharia and Siba Samal for their suggestions, ideas, and critical evaluation of my thesis work during my graduate study.

I would like to acknowledge my lab members for their help and support. I am grateful to Suchitra Derebail for helping me understand various lab techniques and molecular biology concepts when I came to the lab and also for the meaningful discussions we had towards the development of my project. I would also like to thank Megan Heath for helping me out with my experiments during the initial stages. I particularly acknowledge William Bohlayer for taking so many responsibilities in the lab and thank you so much for making the lab website with the primer database. It was very useful. I would like to thank Nirupama Narayanan for her great company and also for making us all laugh with her jokes. I am really going to miss them. I also thank Deena

Jacob for her good company and for all her help during NC preparation. I would like to thank Aminata Mansaray and Nusrat Siddique for making the reagents and for keeping the glassware clean.

I take this opportunity to thank my parents Mary and Anthony for letting me go all the way across the globe with their love and blessings to fulfill my dreams. I also thank my brothers Nikhil and Renil for their encouragement throughout my career. I want to specially thank my dear husband, Rakesh, for his incessant motivation and emotional support. He has always inspired me to excel in whatever I do and has made me a much stronger individual. I cannot thank you enough and cannot wait to spend quality time with you.

Last but not the least, I thank God Almighty for making graduate school a great and memorable learning experience.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
LIST OF ABBREVIATIONS.....	x
Chapter 1 Human Immunodeficiency Virus (HIV); Discovery, Pathogenesis and Life cycle.....	1
1.1 Introduction	1
1.2 Transmission of HIV.....	3
1.3 Origin of HIV.....	4
1.4 AIDS Epidemiology	5
1.5 Pathogenesis of HIV.....	5
1.6 Opportunistic infections and neoplasms in AIDS patients.....	8
1.7 Anti-HIV therapy	9
1.8 HIV structure and genetic organization	10
1.9 HIV life-cycle	16
1.10 Genetic recombination in HIV	22
1.11 Reverse Transcriptase	28
1.12 Nucleocapsid protein	31
1.13 Goals for improving reverse transcription reactions and understanding how RT produces long DNAs in the cell	36
Chapter 2 Formulation of reaction conditions <i>in vitro</i> for the synthesis of long reverse transcription products from genomic RNA segments of Human Immunodeficiency Virus (HIV)	38
2.1 Introduction	38
2.2 Materials	42
2.3 Methods	44
2.4 Results	48
2.5 Discussion	61
Chapter 3 Role of high molecular weight aggregate complexes in the synthesis of long reverse transcription products from genomic RNA segments of HIV.....	65
3.1 Introduction	65
3.2 Materials	68
3.3 Methods	70
3.4 Results	73
3.5 Discussion	80
Chapter 4 Role of dimerization region of HIV and processivity of RT in the synthesis of long reverse transcription products from genomic RNA segments of HIV <i>in vitro</i>	82
4.1 Introduction	82
4.2 Materials	85

4.3	Methods	87
4.4	Results	91
4.5	Discussion	95
Chapter 5 Role of strand transfer in the synthesis of long reverse transcription products from genomic RNA segments of HIV <i>in vitro</i>		
5.1	Introduction	97
5.2	Materials	101
5.3	Methods	103
5.4	Results	110
5.5	Discussion	121
Chapter 6 Characterization of ‘hotspots’ for strand transfer in the C3 env region during Human Immunodeficiency Virus-Type 1 (HIV-1) intersubtype recombination		
6.1	Introduction	124
6.2	Materials	128
6.3	Methods	129
6.4	Results	132
6.5	Discussion	142
Chapter 7 General Discussion		
BIBLIOGRAPHY		155

LIST OF TABLES

Table 1-1: Genes of HIV-1 and function of corresponding proteins	15
Table 5-1: Primers for synthesis of mutations in acceptor	119
Table 5-2: Table showing the nucleotides copied from donor or acceptor at the 5 mutated base positions for eight clones that were sequenced	120

LIST OF FIGURES

Figure 1-1: Human Immunodeficiency Virus	12
Figure 1-2: Dimer Initiation Sequence	13
Figure 1-3: The HIV genome and the proteins encoded by the HIV genes	14
Figure 1-4: Model for HIV replication	21
Figure 1-5: Copy choice model	26
Figure 1-6: Strand displacement-assimilation model	27
Figure 1-7: HIV Reverse Transcriptase	30
Figure 1-8: HIV Nucleocapsid protein	37
Figure 2-1: Schematic representation of the HIV-1 BH10 genome showing length of RNAs and schematic representation of the reverse transcription assay	50
Figure 2-2: Reverse transcription assay with different concentrations of wild type NC using 1.9 kb RNA as template	51
Figure 2-3: Reverse transcription assay with different concentrations of wild type RT using 1.9 kb RNA as template	52
Figure 2-4: An autoradiogram of a time course experiment	54
Figure 2-5: Reverse transcription assay with increasing concentrations of template	55
Figure 2-6: Reverse transcription assay using 4 kb genomic RNA segment as template	56
Figure 2-7: RT titration assay using 4 kb genomic RNA segment as template	57
Figure 2-8: pBR322dR is the pBR322 plasmid having a T7 promoter, HDV ribozyme and T7 terminators	60
Figure 3-1: Reverse transcription assay showing formation of high molecular weight aggregates	74
Figure 3-2: Reverse transcription assay to determine formation of high molecular weight aggregates promoting synthesis of long DNA <i>in vitro</i>	75
Figure 3-3: Reverse transcription reaction to show that inefficient primer-templates binding in reactions with NC	76
Figure 3-4: Reverse transcription assay with increasing concentrations of NC finger mutants	79
Figure 4-1: Reverse transcription assay with increasing NC concentrations using RNA without dimerization signal as template	93
Figure 4-2: Reverse transcription assay to determine role of processivity of RT in the synthesis of long DNA products <i>in vitro</i>	94
Figure 5-1: Reverse transcription assay using RNase H minus RT enzyme (E478>Q)..	112
Figure 5-2: Reverse transcription assay with DNA as template	113
Figure 5-3: Time course strand transfer assay shows that strand transfer is involved in the synthesis of long DNA products	117
Figure 5-4: An autoradiogram of a time course reaction in which excess acceptor was replaced by excess donor	118
Figure 6-1: C1-C4 regions of the HIV <i>env</i> gene and schematic diagram of <i>in vitro</i> template switching assay	135
Figure 6-2: Autoradiogram of template-switching assay from A115 to D89	136
Figure 6-3: Autoradiogram of template-switching assay from A120 to D89	137

Figure 6-4: Autoradiogram of template-switching assay from A115 to A115	138
Figure 6-5: Plots of transfer efficiency vs. time	139
Figure 6-6: Transfer map with regions of strand transfer from A115 donor to D89 acceptor and table showing the number of recovered transfer products ...	140
Figure 6-7: Predicted structure of RNAs from clones A115, A120, and D89 over bases 7255-7454 (relative to HIV HXB2 provirus)	141
Figure 7-1: Proposed model for synthesis of long DNA products in vitro	148, 149

LIST OF ABBREVIATIONS

HIV	Human Immunodeficiency virus
AIDS	Acquired Immunodeficiency syndrome
ERT	Endogenous reverse transcription
NERT	Natural endogenous reverse transcription
LTR	Long terminal repeat
RT	Reverse Transcriptase
RNase H	Ribonuclease H
PBS	Primer binding site
PPT	Poly-purine tract
AMP	Adenosine mononucleoside phosphate
rNTPs	Ribonucleoside triphosphates
dNTPs	Deoxyribonucleotide triphosphates
DTT	Dithiothreitol
PNK	Polynucleotide kinase
EDTA	Ethylenediaminetetraacetic acid
PCR	Polymerase chain reaction
ss	Single-stranded
ds	Double-stranded
IN	Integrase enzyme
PR	Protease enzyme
NC	Nucleocapsid protein
TAR	<i>Trans</i> -activation response

RRE	Rev response element
kb	Kilobase
nt	Nucleotides
cDNA	Complementary DNA

Chapter 1 Human Immunodeficiency Virus (HIV); Discovery, Pathogenesis and Life Cycle

1.1 Introduction

Human Immunodeficiency Virus (HIV) is a member of the lentivirus genus (Latin, *lentus*: slow, pertaining to slow growth) of the family *Retroviridae* (Retroviruses) which comprises a large and diverse group of enveloped RNA viruses (1, 2). Retroviruses were so named because their growth cycle includes a very important step of copying of the genomic RNA into DNA by a virus-coded polymerase called 'reverse transcriptase', thereby reversing the flow of genetic information. This was an exception to the then-accepted 'Central Dogma' of Molecular Biology that the flow of genetic information is unidirectional from DNA to RNA to protein. HIV (Human Immunodeficiency Virus) is the causative agent of Acquired Immune Deficiency Syndrome (AIDS). AIDS is called a syndrome as it is a collection of about 27 known diseases (mostly opportunistic infections) and symptoms resulting from severe impairment of the immune system. According to the Centers for Disease Control and Prevention, Atlanta, Georgia, a person is diagnosed with AIDS, if he or she tests positive for antibodies to HIV and has a T4 lymphocyte count below 200/ μ l of blood or has any of the 27 diseases (3) .

HIV was first isolated in January 1983 from the lymph node of a patient with lymphadenopathy by Françoise Barré-Sinoussi with a group headed by Luc Montagnier in the laboratory of Jean-Claude Chermann at the Pasteur Institute in Paris (4). They

named the virus “lymphadenopathy-associated virus” (LAV) (5). It was also isolated by Robert Gallo and coworkers at the U.S National Institutes of Health and named Human T-Cell Lymphotropic Virus III (6). At around the same time it was also isolated by Jay Levy and coworkers at the University of California, San Francisco and named AIDS-related virus (ARV) (7). Electron microscopic studies revealed that all these virus isolates were morphologically similar and hence in 1986, the Human Retrovirus Sub-committee of the International Committee on the Taxonomy of Viruses coined its current name: human immunodeficiency virus which has now been accepted worldwide. In 1986, François Clavel and coworkers isolated a novel type of HIV that is prevalent in some parts of West Africa and was called HIV-2 to distinguish it from the original isolate, HIV-1 (8). HIV-1 and HIV-2 are the only known human Lentiviruses. Of the two, HIV-1 is predominant and found throughout the world, whereas HIV-2 has been isolated primarily in West Africa with some cases in America and Europe. Individuals infected with HIV-2 also develop AIDS but studies suggest that the incubation period for HIV-2 for the development of progression of disease is longer than for HIV-1 (9) and HIV-2 is not as easily transmitted perinatally as HIV-1 (10). HIV-1 has been categorized into three groups; Group M (for main) is comprised of viruses responsible for the majority of infections throughout the world, Group O (for outlier), a rare group found in Cameroon, Gabon and France and Group N (new group) was found in AIDS patients from Cameroon. Group M is further divided into ten subtypes or clades (A through K) (11, 12). All work done in this dissertation report applies to HIV-1 although similarities between the two viruses suggest that the general conclusions would hold for both types.

1.2 Transmission of HIV

HIV is transmitted by sexual contact through vaginal, rectal and penile tissues; by direct injection with HIV-contaminated drugs, needles, syringes, blood or blood products; and from an HIV-infected mother to her fetus in *utero*, from a mother to her infant during childbirth or via breast milk (13). There is no evidence that the virus can be transmitted by casual contact or by insect vectors. The predominant mode of transmission in the United States and Western Europe is by homosexual contact and by sharing of contaminated needles during drug use. A large majority of infections worldwide, especially in the developing countries is via heterosexual contact. HIV can be transmitted to either partner during vaginal intercourse (14) with a probability of a woman becoming infected by her HIV-positive male partner during vaginal intercourse being less than 0.2% and the risk of infection from a woman to a man during vaginal intercourse is even less likely (15). The presence of ulcerative genital lesions due to other sexually transmitted diseases increases the susceptibility to HIV infection. This may be due to a greater chance of HIV entering the bloodstream through such a lesion or a higher incidence of T4 target cells in the lesion. Of the routes for sexual transmission, receptive anal intercourse appears to have the highest risk per incident rate. Before HIV was identified as the causative agent of AIDS, it was accidentally transmitted by contaminated blood and blood products. Screening of the blood supply and preventing recognized at-risk individuals from donating blood has significantly reduced the number of cases infected via this mode. The risk of infection of a health care worker from an infected individual via a sharp object such as a needle stick is approximately 1 in 300 (13). More than 90% of HIV infections in infants and children are by vertical

transmission. The rate of mother-to-infant transmission of HIV in absence of intervention range from 13%- 42%. HIV may be transmitted when maternal blood enters fetal circulation or by mucosal exposure to HIV during labor and delivery. Breast-feeding adds an estimated 14% additional risk over the risk of intrapartum or perinatal HIV infections. A specific regimen of zidovudine (AZT) has been found to reduce the risk of perinatal transmission of HIV (16).

1.3 Origin of HIV

There are two possibilities for the origin of HIV. One possibility is that the virus has always existed in isolated populations and may not have been recognized due to low prevalence or geographical confinement. Modern society, urbanization and extensive world travel would have made the virus recognizable on a global front. Another possibility is that HIV-1 may have recently entered the human population from another species. The most likely origin may be a combination of the above two possibilities. Which possibility is correct may never be resolved. Recent evidence shows that all HIV-1 strains known to infect man, including HIV-1 groups, M, N and O are closely related to one SIVcpz lineage found in *Pan troglodytes troglodytes*, a chimpanzee subspecies in Africa and *Pan troglodytes troglodytes* is the primary natural reservoir for HIV-1 (17, 18). The date of the last common ancestor of the main group of HIV-1 is estimated to be 1931 (19). HIV-2 infection in humans in western Africa may have emerged or may be still occurring by cross-species transmission from sooty mangabey monkeys (20). The forested regions of western Africa which is the natural habitat of these monkeys almost coincides with the region where HIV-2 human infection is endemic and the sequences of

HIV-2 isolates are within the range of variation of known SIV isolates infecting sooty mangabey monkeys (21).

1.4 AIDS Epidemiology

Since the first clinical evidence of HIV/AIDS in 1981, AIDS has killed more than 25 million people (22). The epidemic claimed 3.1 million lives in 2005 alone, more than half a million were children. Nearly 5 million people were newly infected with the virus in 2005 making the total number of people living today with HIV reach its highest level with an estimated 40.3 million people infected with HIV, double the number (19.9 million) in 1995 (22). Sub-Saharan Africa continues to be the hardest hit with approximately 25.8 million people infected with HIV. There is an increase in the proportion of women being infected with HIV. In 2005, 17.5 million women were living with the virus which is one million more than in 2003. Approximately 2.3 million children under 15 years of age were living with the virus in 2005 with 700,000 new infections in 2005 alone (22). Prevention of HIV in healthy individuals and providing health care and affordable treatment presents one of the biggest challenges facing us today.

1.5 Pathogenesis of HIV

HIV enters the body via infected body fluids like blood, semen and vaginal secretions. Once inside the body, HIV attacks a subset of lymphocytes called T4 lymphocytes. The T4 lymphocytes are called Helper T cells. These cells by means of their T cell antigen receptor (TCR) recognizes antigenic peptides in association with self

major-histocompatibility-complex (MHC) molecules displayed on the surface of antigen-presenting cells (APC) like B cells or macrophages, interact with killer T cells and B cells and help them respond to foreign antigens. The primary receptor for HIV on the T4 cell is CD4 (23, 24). T4 cells carry about 10,000 copies of CD4 molecules on their surface. CD4 molecule serves as an adhesion molecule that stabilizes the interaction between the APC displaying the antigenic peptide and the TCR (25, 26). In addition to the primary receptor, a co-receptor is also required for HIV entry into the host cell (27, 28, 29). These co-receptors normally function as chemokine receptors and are believed to be involved in migration of cells to areas producing chemokines (small secreted molecules that serve as chemo-attractants). Different variants of HIV-1 can use different chemokine receptors with the fusin (CXCR4) and CCR5 receptors being the most common. Chemokine receptors and CD4 molecules are also present on macrophages which are also productively infected by HIV. Upon entry, HIV destroys the T4 cell in various ways. The copious budding of HIV from the T4 cell surface results in disruption of integrity of the cell membrane. Cellular RNA and protein synthesis is also disrupted due to high levels of viral RNA and unintegrated viral DNA into the cell (30). The binding of gp120 (viral envelope protein) to CD4 causes the T4 cells to lose their immune function and can cause them to become targets for immune attack by antibody mediated antibody-dependent cell cytotoxicity (ADCC). Formation of syncytia or multinucleated giant cells is one of the causes of death of cells that are not directly infected with HIV (31, 32). During HIV infection, cross-linking of CD4 molecules by HIV gp120 or gp120-anti gp120 immune complexes prepares the cell for apoptosis (33, 34). The activation of a prepared cell by an antigen or superantigen could lead to death of the cell by apoptosis without direct

infection with HIV (35). Uninfected cells may also be destroyed by autoimmune mechanisms. Antigenic cross-reactivity exists between HIV proteins (gp120 and gp41) and MHC class I and II determinants (31, 36), which causes anti-HIV antibodies to eliminate uninfected MHC class I and II bearing cells via ADCC. The decline in T4 cells that occurs during viral infection may be in part due to the inability of the immune system to regenerate the T4 cell pool (37). This may be due to the destruction of lymphoid precursor cells by infection as they get activated to divide and also due to disruption of the microenvironment required for efficient regeneration of immune competent cells. A healthy individual has T4 cell counts between 800 to 1200/ μ l of blood. Immediately following initial infection by HIV, there is a steep decline in the number of T4 cells with subsequent recovery to near normal level. This is followed by an average yearly loss of about 60 T4 cells/ μ l (38). An HIV positive person is diagnosed with AIDS when the T4 counts drop below 200/ μ l. This state is also associated with profound immunodeficiency and multiple or disseminated opportunistic infections (39). In general, this would take from 8 to 12 years from initial infection in the early years of the AIDS epidemic (40-42). Better detection of HIV infection, antiretroviral therapy and effective prophylaxis for opportunistic infections today has considerably slowed down progression of HIV disease and has increased life expectancy in HIV-positive individuals in those areas where treatment is available (43, 44).

1.6 Opportunistic infections (OIs) and neoplasms in AIDS patients

Primary infection with HIV results in acute mononucleosis-like clinical syndrome approximately 3-6 weeks following infection. This syndrome is accompanied by a burst of viral replication that can be detected in blood approximately 3 weeks following infection (45, 46). An antiviral immune response is detected approximately 3-6 weeks after infection (47, 48). Following the induction of an immune response, there is a very long period of clinical latency characterized by very few and mild, if any, clinical manifestations. When T4 cell counts drop below 500/ μ l, usually the first symptoms appear in the HIV-positive individual. When the counts fall below 200/ μ l, the patient is susceptible to AIDS-defining opportunistic infections (OIs) and neoplasms. About 90% of AIDS related deaths are caused by OIs, 7% are due to neoplasms and 3% due to other causes. A wide range of bacteria, fungi, protozoans and viruses, most of which are harmless commensals in a healthy individual become opportunistic pathogens in an AIDS patient due to a failing immune system. Some of the common OIs associated with AIDS are candidiasis, *Pneumocystis carinii* pneumonia (PCP) and cytomegalovirus infections. About 40% of AIDS patients develop neoplasms like Burkitt's lymphoma, Kaposi's sarcoma, B cell lymphomas and cervical carcinomas. Various neurological symptoms like AIDS dementia complex (ADC), aseptic meningitis and various myelopathies are also seen in some full blown AIDS patients (13).

1.7 Anti-HIV therapy

There are four main classes of anti-HIV drugs. They are:

- Nucleoside analog RT inhibitors (NRTIs) eg. Zidovudine, Didanosine, Lamivudine. These nucleoside analogs prevent HIV replication by their incorporation into the elongating strand of viral DNA thereby causing chain termination.
- Non-nucleoside analog RT inhibitors (NNRTIs) eg. Nevirapine, Delvaridine. These drugs inhibit HIV replication directly by binding noncompetitively to reverse transcriptase enzyme.
- Protease inhibitors eg. Saquinavir, Ritonavir, Indinavir. These drugs are made up of a small number of amino acids (up to 15) or amino acid analogs and bind to the active site of the viral protease enzyme and inhibit its activity. This inhibition prevents cleavage of long HIV protein precursors, resulting in the formation of immature noninfectious HIV particles.
- Fusion inhibitor eg. Enfuvirtide (Fuzeon). Enfuvirtide was approved by the U.S. FDA in March 2003. Enfuvirtide is approved for HIV patients who have tried all the currently available antiretrovirals but have failed to keep their viral loads to undetectable levels. It is not approved for people who are starting anti-HIV drugs for the first time. When administered it must be used in combination with other drugs. Enfuvirtide belongs to a category of medications called fusion inhibitors or entry inhibitors. It is a peptide which binds to gp41 protein on HIV's surface. Once it does this, HIV cannot successfully bind with the surface of T-cells, thus preventing the virus from infecting healthy cells (49).

All the above drugs are administered as multi-drug cocktails to overcome the rapid emergence of drug resistant mutants. This combination therapy is called HAART (Highly Active Anti-Retroviral Therapy). Although in past years patients were typically started on HAART therapy immediately after HIV diagnosis, current protocols suggest waiting until T helper counts are at or below 350 cells/ μ l. AIDSVAX, developed by a company in Brisbane, California, is the only FDA approved vaccine to have reached phase III trials (50). This vaccine uses recombinant gp120, an envelope glycoprotein. However, the vaccine failed to prevent HIV infection in the study population after three years of clinical trial (51). There are several experimental recombinant vaccines produced by various companies that are undergoing Phase I and Phase II trials (52).

1.8 HIV structure and genetic organization

HIV is about 80-100 nm in diameter (see Figure 1-1). The outer lipid layer of the virus is derived from the cell lipid bilayer into which glycoproteins (gp120 and gp41) encoded by the *env* region of the viral genome are incorporated. Internal to the envelope is the matrix made up of the MA protein (for membrane associated or matrix). The matrix encapsidates the shell referred to as capsid and made up of the CA protein (for capsid). The capsid together with the components it encloses is referred to as the “core” which is cone-shaped (53). Within the core is the RNA genome which is approximately 9.7 kb and is linear, single-stranded, non-segmented and of positive polarity (53). There are two completely identical or nearly identical copies of the RNA genome non-covalently linked in an apparent parallel orientation close to their 5' ends by a structure called the Dimer Linkage Structure (DLS) or the Dimer Initiation sequence (DIS) (Figure 1-2). Although

there are two copies, generally only one provirus (virus encoded dsDNA that is integrated into the host genome) is detected in a single infection, hence retroviruses are considered to be pseudodiploid (54). The existence of two genomes probably helps retroviruses survive and complete replication in case of damage to any or both of the genomes and may also account for the high rate of genetic recombination seen in these viruses (54, and see below). The genome RNA is coated along its length with a viral nucleocapsid protein (NC). The HIV genome contains nine genes, which produce at least nine proteins (see Figure 1-3). These proteins are classified into: three structural polyproteins encoded by *gag*, *env* and *pol* genes, two regulatory proteins encoded by *tat* and *rev* genes and four accessory proteins encoded by *nef*, *vif*, *vpr* and *vpu*. The *gag* (for group-specific antigen) gene codes for the above mentioned MA, CA and NC proteins. The *pol* gene codes for the viral enzymes RT (Reverse Transcriptase), PR (Protease) and IN (Integrase) while *env* codes for the two glycoproteins also mentioned above: gp120 which forms the external “spikes” of the virus and gp41 which forms the transmembrane protein that connects gp120 to the viral capsid surface. Table 1-1 lists the nine genes, their proteins and functions.

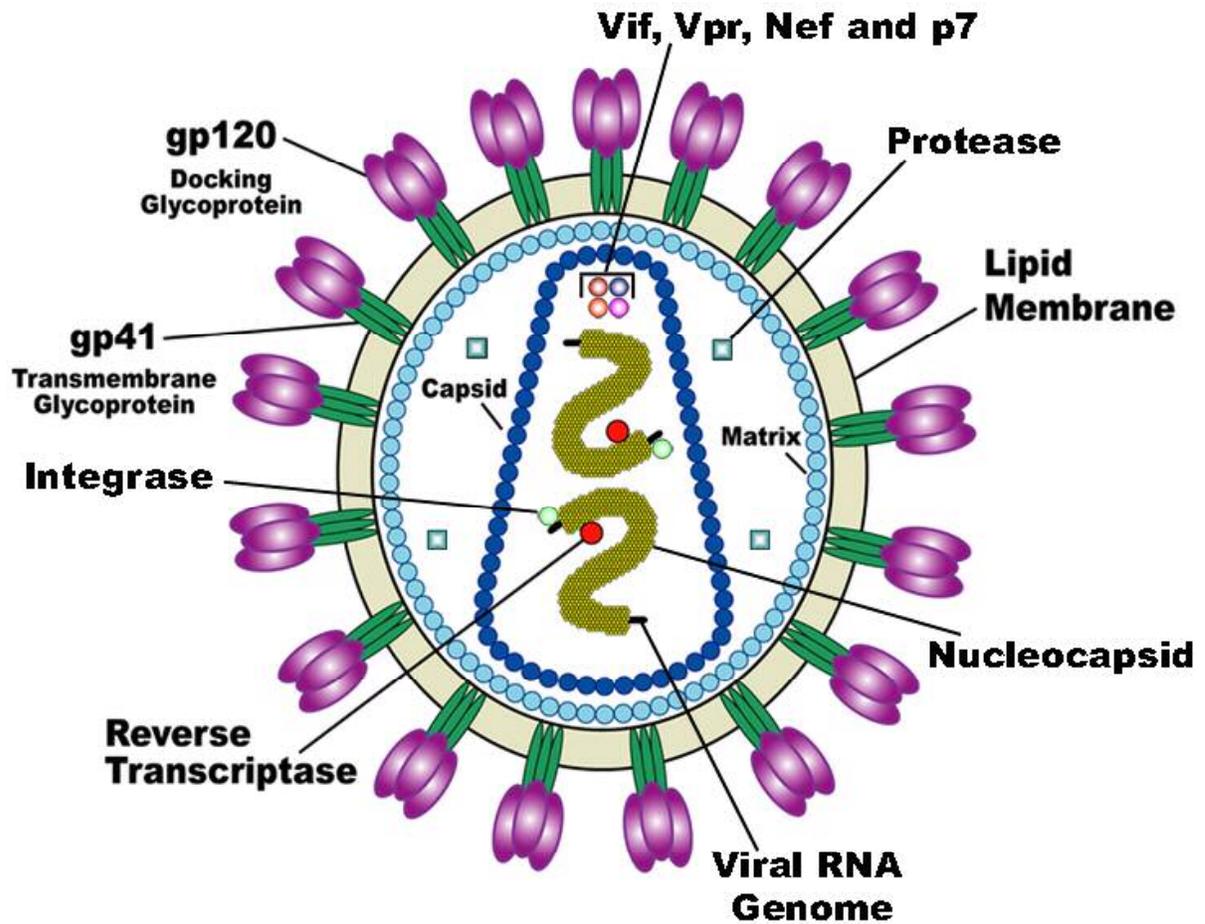


Figure 1-1: Human Immunodeficiency Virus

Shown is a schematic of HIV. It is an enveloped virus containing glycoproteins gp 120 and gp 41. Within the envelope are the matrix protein (MA) and the capsid (CA) protein. The capsid encloses two copies of single-stranded plus sense RNA genomes coated with nucleocapsid protein. The capsid also contains several copies of reverse transcriptase, integrase and protease enzymes. Figure adapted from [www.en.wikipedia.org](http://en.wikipedia.org/wiki/HIV_structure_and_genome)
http://en.wikipedia.org/wiki/HIV_structure_and_genome
http://en.wikipedia.org/wiki/HIV_structure_and_genome

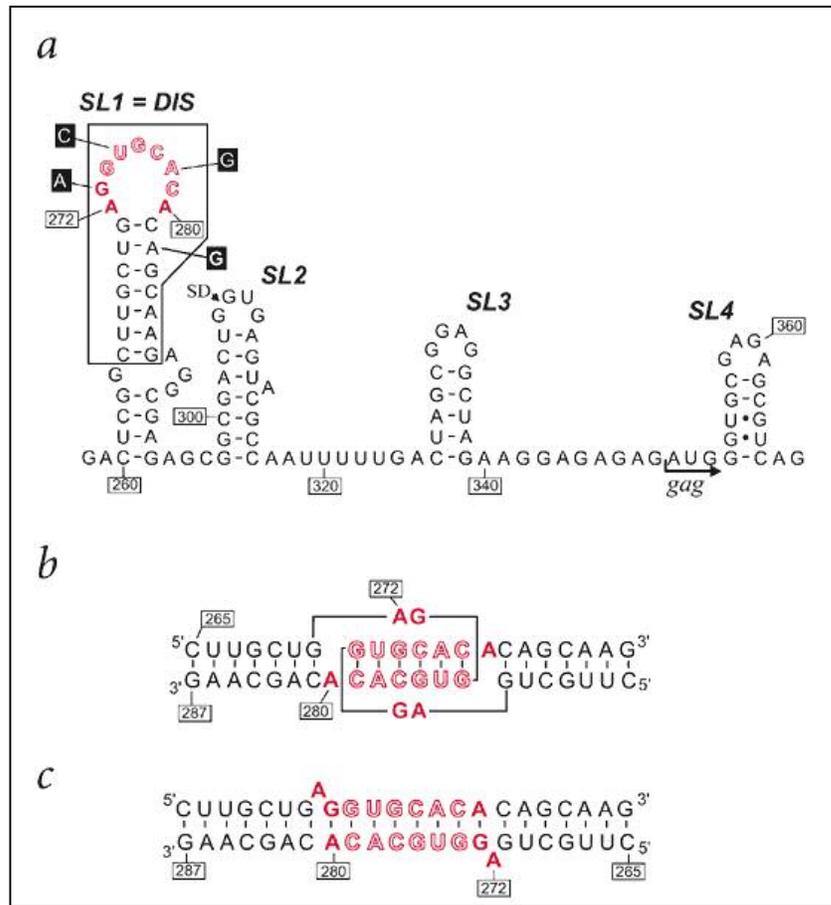


Figure 1-2: Dimer Initiation Sequence

Localization and structure of Dimer Initiation Sequence (DIS) of HIV-1. Bases that constitute the kissing loop as well as the flanking purines are in red.

- shows the secondary structure of the packaging signal consisting of the four stem loops, SL1, SL2, SL3 and SL4.
- Secondary structure of the DIS dimer forming a kissing-loop complex.
- Secondary structure of the extended duplex

Figure taken from Ennifar E. *et al*, 2001. *Nature Structural Biology* **8**: p. 1064-1068

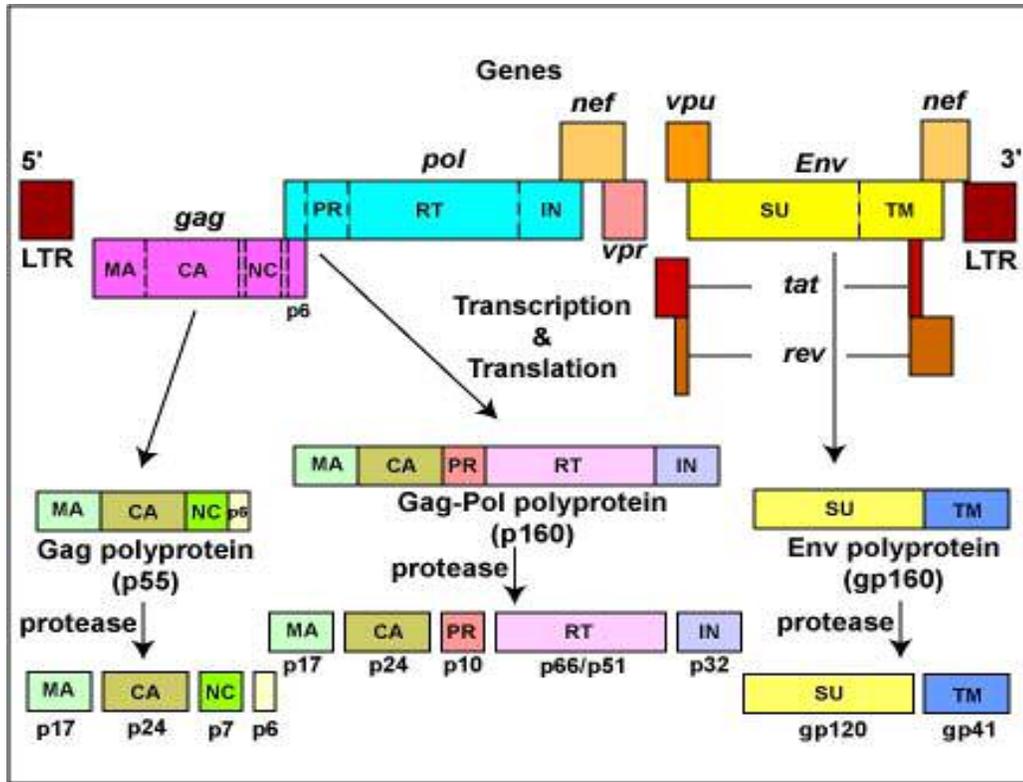


Figure 1-3: The HIV genome and the proteins encoded by the HIV genes

Shown are the HIV genome and the proteins encoded by the genes. The nine genes, their relative positions, and the positions of the LTRs or long terminal repeats are shown. The products resulting from *gag*, *pol* and *env* are described. *Tat* and *rev* are both shown upstream and downstream of *env* since splicing within the *env* gene is required to complete the coding of the *tat* and *rev* proteins. Figure adapted from www.cat.cc.md.us

Table 1-1: Genes of HIV-1 and function of corresponding proteins

Gene name(s)	Mol. Mass (kDa) ¹	Function
<i>gag</i> (group Antigen)	p17 p24 p7	Matrix (MA, interphase between capsid and envelope, virion structural protein) Capsid (CA, protection, virion structural protein) Nucleocapsid (NC, protection, packaging, dimerization, reverse transcription)
<i>pol</i> (polymerase)	p10 p66/ p51 p31	Protease (PR, cleavage of polyproteins) Reverse transcriptase (RT, converts viral RNA into double-stranded DNA, functions as a p66/p51 heterodimer with the latter being the N-terminal cleavage product of p66) Integrase (IN, integrates proviral DNA into host cell DNA)
<i>env</i> (envelope)	gp120 gp41	Surface glycoprotein binds to receptor molecule during viral entry Transmembrane glycoprotein fuses with cell membrane during viral entry
<i>tat</i> (transactivator protein)	p14	Stimulates viral transcription by binding to TAR ² region to facilitate initiation and elongation of viral transcription
<i>rev</i> (differential regulator of expression of virus protein)	p13/ p19	Binds to RRE ³ to facilitate nuclear export of unspliced or singly spliced RNA and increases production of structural proteins
<i>vif</i> (virus infectivity factor)	p23	Increases virus infectivity, affects virion assembly and viral DNA synthesis and suppresses inhibition of replication by human anti-HIV component APOBEC3G
<i>nef</i> (negative regulatory factor)	p27	Increases or decreases virus replication, affects T-cell activation and enhances virus infectivity
<i>vpr</i> (virus protein R)	p15	Causes G2 arrest, aids in the nuclear transport of pre-integration complexes
<i>vpu</i> (virus protein U)	p16	Destroys CD4 on the surface of infected cell thereby allowing budding of virions

¹Table I lists the size of the functional proteins

²TAR or the trans-activation response sequence is a stable, stem-loop present at the 5' end of all HIV RNAs. It is highly conserved and required for Tat function

³RRE or Rev response element is a highly structured 351 nt coding region within the *env* region. It is present in RNAs that are dependent on Rev protein for their expression in the host cytoplasm

1.9 HIV life cycle

Viral entry: The process of entry of HIV into the target cell is the first step in the infection cycle. It is initiated through the binding of viral surface glycoprotein (gp120) to CD4 receptor molecule on the outer membrane of the target cell. Gp120 together with a membrane-spanning protein (gp41) constitute the envelope glycoprotein complex. In addition to CD4 receptor molecule, viral entry requires a second receptor, β -chemokine receptor called "FUSIN" or CXCR4 present on T cells and CCR5 present on macrophages. The binding of gp120 to CD4 molecule apparently triggers a conformational change in the glycoprotein (55) which activates the membrane fusion-inducing potential of gp41 (56) which then mediates fusion of the lipid bilayers of the cell and viral membranes (57). Initially, it was shown that HIV enters the cell by receptor-mediated endocytosis (58), later on, it was shown that entry is via a pH-independent method (59, 60). For fusion to occur adjacent membranes merge and a fusion pore is formed but not much is known about these steps and this is an area of current investigation.

Reverse transcription: The process of reverse transcription was studied by identifying replication intermediates synthesized in reconstituted or endogenous reactions and also by analysis of viral DNA isolated from infected cells (61-64). Reverse transcription begins when the viral particle enters the cytoplasm of the host cell. The viral genome enters the cytoplasm in the form of a ribonucleoprotein complex that also has reverse transcriptase (RT), nucleocapsid protein (NC) and presumably Vpr (65) The exact nature of the replication complex is unclear and its form presumably changes as reverse

transcription proceeds. RT is a multifunctional enzyme that has RNA directed and DNA directed polymerase activity and RNase H activity (ribonuclease H) that cleaves the RNA strand that is part of an RNA:DNA duplex. NC is a viral protein that has numerous roles (see below) and is known to increase the efficiency of reverse transcription. NC is a nucleic acid chaperone protein (66, 67). Like other chaperones NC presumably binds nucleic acids and aids in proper folding by preventing misfolding and resolving incorrectly folded structures (see below). Although the genomic RNA of HIV contains immediately translatable information, the virus during its life cycle goes through an intermediary double-stranded DNA stage *via* reverse transcription (68) as shown in Figure 1-4.

The HIV RNA is derived from the double-stranded integrated viral DNA by the RNA synthesis machinery of the host and bears the structural features of a cellular mRNA in that it has a 7-methylated-G cap and a polyadenylated 3' end. Immediately adjacent to the 5' cap and the 3' poly (A) tail at the 5' and 3' end respectively, lie the direct repeats, termed R for repeat sequences (69). Adjacent and internal to the R sequence on each end are sequences that are unique to that end called U5 (unique to 5' end) and U3 (unique to 3' end). PBS is the primer binding site for tRNA primer. The tRNA primer and PPT (the polypurine tract) serve as RNA primers for minus and plus strand DNA synthesis respectively. The *gag*, *pol* and *env* genes are also shown in Figure 1-3. The overall arrangement of these different sequences on the HIV genome is 5' cap, R, U5, PBS, *gag*, *pol*, *env*, PPT, U3, R, poly (A). LTR or long terminal repeats at both

ends are seen only in the proviral DNA (Figure 1-4). These are formed as a result of duplication of U5 and U3 sequences. The generation of the two LTRs at both ends of the proviral DNA is essential to establish DNA that can be integrated into the host chromosome and to regenerate sequences (the promoter and polyadenylation sites), which are essential for viral transcription.

Following are the steps of reverse transcription:

1. The process of reverse transcription is initiated from a host-derived tRNA^{Lys3} that is packaged in the virion during assembly. The tRNA partially unwinds and the first 18 bases at the 3' end bind to the complementary primer binding site (PBS) near the 5' end of the genomic RNA. The process is aided by NC protein (70). The RT enzyme using its RNA dependent DNA polymerase activity synthesizes the minus strand DNA until the 5' end of the genomic RNA is reached generating what is referred to as minus-strand strong-stop DNA. The minus-strand strong-stop DNA (-sss DNA) is about 150 bases long.
2. When the minus strand strong-stop DNA is formed, the RNase H activity of RT degrades the RNA strand of the RNA:-sssDNA hybrid. This releases the DNA and allows the first strand transfer (also referred to as template jumping or template switching) to occur and the -sssDNA binds to the complementary repeat (R) region at the 3' end of the same (intramolecular) or the second genomic RNA (intermolecular) (71). This process is also aided by NC (72).
3. The -sss DNA annealed near the 3' end of the template is extended up to the PBS on the RNA template. The RNA template is simultaneously degraded by RT's RNase H activity. The RNase H cleavage products include a polypurine tract or PPT that serves as

primer for synthesis of plus strand DNA (73). The PPT is a highly conserved 15 base oligomer that is resistant to degradation by RNase H activity of RT.

4. RT uses the DNA dependent DNA polymerase activity to synthesize the second strand from the 3' end of the PPT using the minus strand DNA as template until it reaches the first modified base (19th base from the 3' end of the tRNA) on the tRNA primer that cannot serve as template for reverse transcription (74). The resultant DNA product is called plus strand strong-stop DNA or +sssDNA.

5. The tRNA at this point is removed by RNase H activity of RT, which allows the second strand transfer to occur involving the +sssDNA. At this point, the minus strand of the PBS is copied from the viral RNA and the plus strand of the PBS sequence is copied from the tRNA primer. In the second strand transfer, the PBS sequence near the 3' end of the +sssDNA binds to the PBS sequence near the 3' end of the minus strand DNA. This transfer is intramolecular and leads to the formation of a circular DNA molecule with overlapping 5' ends. Synthesis of both the plus and minus strand DNAs are completed to produce a linear double stranded viral DNA molecule with long terminal repeats (LTRs) (75).

This molecule then integrates into the host chromosome in a reaction catalyzed by integrase enzyme (76). Integration is random although sites of active transcription appear to be favored (77). Integrase which is proteolytically cleaved from the carboxy-terminal portion of Pol polyproteins recognizes specific target sequences that are approximately 15 nucleotides in length at the ends of the linear proviral DNA and promotes their integration into the host chromosomal DNA (78). Once integrated (now referred to as a

provirus), the viral DNA is stable and the virus now utilizes the host RNA synthesis machinery for transcription of viral DNA into new copies of the viral genome and also mRNAs that encode viral proteins. The proteins are assembled (some as precursors and others in mature forms) and packaged into virions along with two copies of RNA genome. The new viral particles are released from the cell by budding. After the budding process protein precursors are cleaved by PR and the core structure is formed, generating infectious viral progeny.

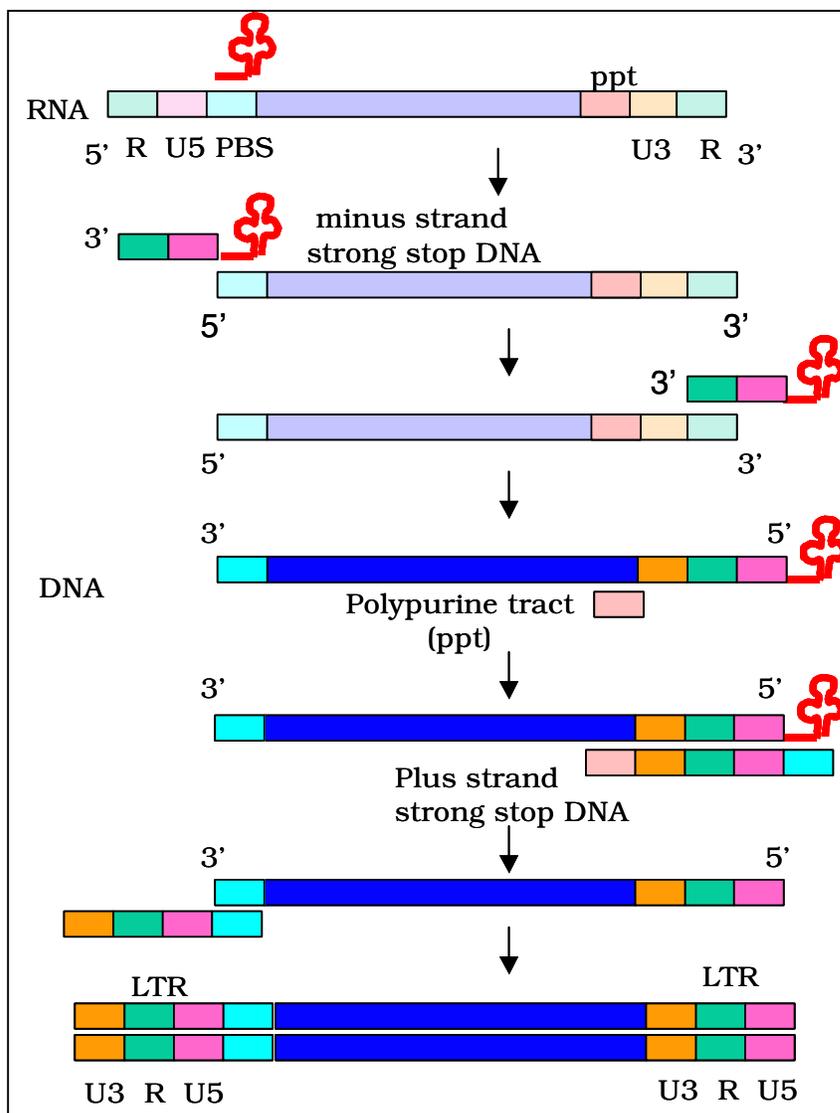


Figure 1-4: Model for HIV replication Shown here is the model for retroviral replication. The synthesis of proviral DNA begins when a tRNA primer binds to the primer binding site (PBS). Synthesis of this DNA extends to the 5' end of the viral genome. This segment of DNA is called 'minus strand strong-stop DNA' (-sssDNA). The -sssDNA undergoes transfer from the 5' end to the 3' end of the same or different RNA genome. After transfer, synthesis continues to the 5' end of the RNA. At the same time, the RNA gets cleaved by the RNase H activity of RT. The ppt or polypurine tract is the RNase H resistant segment of the genome that primes the plus strand DNA synthesis. This phase of synthesis produces a segment called 'plus strand strong-stop DNA' (+sssDNA). The second transfer involves binding of the complementary PBS regions of minus and plus DNA strands. Complete synthesis yields a double stranded DNA with long terminal repeats (LTRs).

1.10 Genetic recombination in HIV

Recombination occurs at a high rate during the retroviral replication cycle. Recombination is one of the means by which the virus generates genetic diversity. Because of this genetic diversity, HIV-1 has been defined as a quasispecies which is a population of closely related yet genetically distinct viruses within the same infected individual (79). Retroviruses including HIV copackage two RNA genomes. A consequence of this packaging is a high rate of recombination (54). These genomes are identical or genetically distinct. Genetically distinct genomes can arise if a single cell has been coinfecting by two different viruses (80) resulting in more than one provirus in which case the RNA genomes could be generated from different proviruses and copackaged into the same virion. Also, the viral RNA genomes are transcribed from the integrated provirus by host RNA polymerase II which could introduce some errors.. This can give rise to genetically distinct RNA genomes that may be copackaged into the same virion. Recombination in retroviruses is mechanistically different from what occurs in higher life forms. Recombination in retroviruses occurs during synthesis of the double-stranded DNA from the genomic RNA. It occurs by a process called strand transfer (also referred to as ‘strand jumping’ or ‘template switching’). Strand transfer involves the switching of DNA being synthesized on one template (referred to as ‘donor’) to homologous regions on the same or on a second template (referred to as ‘acceptor’) where synthesis of DNA continues. When strand transfer occurs on to a genetically distinct template, the proviral DNA so obtained is a chimera of the original parent templates. The first and second strand transfers (called minus strand and plus strand strong stop DNA transfers respectively) are obligatory strand transfers, which occur at

the termini of genomic RNA during reverse transcriptase (described above). These transfers are essential steps without which viral replication cannot proceed to completion. In addition to these obligatory transfers, internal transfers can take place potentially at any position on the genome (81, 82). Such internal strand transfers may increase the probability of successful DNA synthesis by providing a salvage pathway for damaged or broken RNA genomes (83). These internal transfers also help in generating genetic diversity in the population (84-86) thereby allowing viruses to escape the host immune response and evade drug therapy. Internal strand transfer can occur during minus and plus strand DNA synthesis (84). However, these transfers occur to a lesser extent during plus-strand DNA synthesis compared to minus strand DNA synthesis (71). Internal strand transfers occur between homologous regions of two genomic RNAs or DNAs synthesized from those RNAs (54). Non-homologous recombination is very rare occurring at a rate of about $1/100^{\text{th}}$ to $1/1000^{\text{th}}$ of the rate of homologous recombination (87). It has been shown that HIV-1 recombines approximately two to three times in every replication cycle (81). More recent results indicate that the virus may be more recombinogenic with an average of nine recombination events per virus in T lymphocytes and about thirty crossover events in macrophages (88). These very high rates for HIV are in contrast to some simpler retroviruses like Moloney murine leukemia virus and spleen necrosis virus that show rates about $1/10$ to $1/50^{\text{th}}$ the rate for HIV (54, 89).

Two models have been proposed for the mechanism of retroviral recombination: copy-choice model and strand-displacement assimilation model (90). There is experimental evidence for both and they are not mutually exclusive. The copy-choice

model explains recombination during minus-strand DNA synthesis and the strand-displacement assimilation model explains recombination during plus-strand DNA synthesis. The copy-choice model (Figure 1-5) postulates that recombination occurs when the growing DNA molecule switches from one RNA template to another during minus-strand DNA synthesis. The copy-choice model is a modified version of the original 'forced copy-choice' model. The forced copy-choice model proposed that strand transfer occurs when there is a break or damage in the template RNA and this forces the growing strand to a homologous region of the copackaged RNA (83). However, this model did not explain all types of transfers during minus strand synthesis. For example, Hu and Temin (91) introduced breaks into genomic RNA by gamma radiation but this did not significantly enhance recombination. Template switching has been shown to occur efficiently from regions of unbroken RNA (92). In undamaged or broken RNA, secondary structures and/or sequences were shown to be responsible for pausing or stalling of RT which could result in strand transfer (93, 94). Misincorporation by RT is also known to induce strand transfer (95). Thus, the modified version called copy-choice was proposed to account for all types of transfers that occur during minus strand DNA synthesis. The transfer of the growing DNA molecule can occur by two proposed methods (96): DNA dissociation and acceptor RNA invasion. In the DNA dissociation method, the growing DNA dissociates from the original template (called donor) before annealing to the second template (called acceptor). This DNA dissociation is independent of the acceptor template. In the acceptor RNA invasion method, the acceptor RNA actively displaces the growing DNA from the donor template by invading the donor

RNA-growing DNA hybrid. During this invasion, a trimeric structure is formed transiently with the DNA bound to both donor and acceptor templates. The existence of such a transient trimeric structure has been shown both *in vitro* and in cell culture (95, 97). There is evidence for transfer by DNA dissociation method also. It is however not known whether transfer during minus strand DNA synthesis occurs by one or both these methods.

The strand-displacement assimilation model (Figure 1-6) explains strand transfer occurring during plus strand DNA synthesis (98). Plus strand DNA synthesis originates primarily from the polypurine tract although it can be initiated at multiple alternative points (73). Therefore plus strand DNA synthesis is discontinuous. The strand-displacement assimilation model postulates that during such a discontinuous synthesis, the 5' ends of the growing plus strand DNA are displaced by the 3' ends of adjacent plus strands. These strands are now free and can base pair with another minus strand DNA. This model requires the presence of two minus strand DNA molecules containing the region where recombination will occur. Evidence also suggests that concomitant DNA synthesis may be required for strand displacement to occur (99).

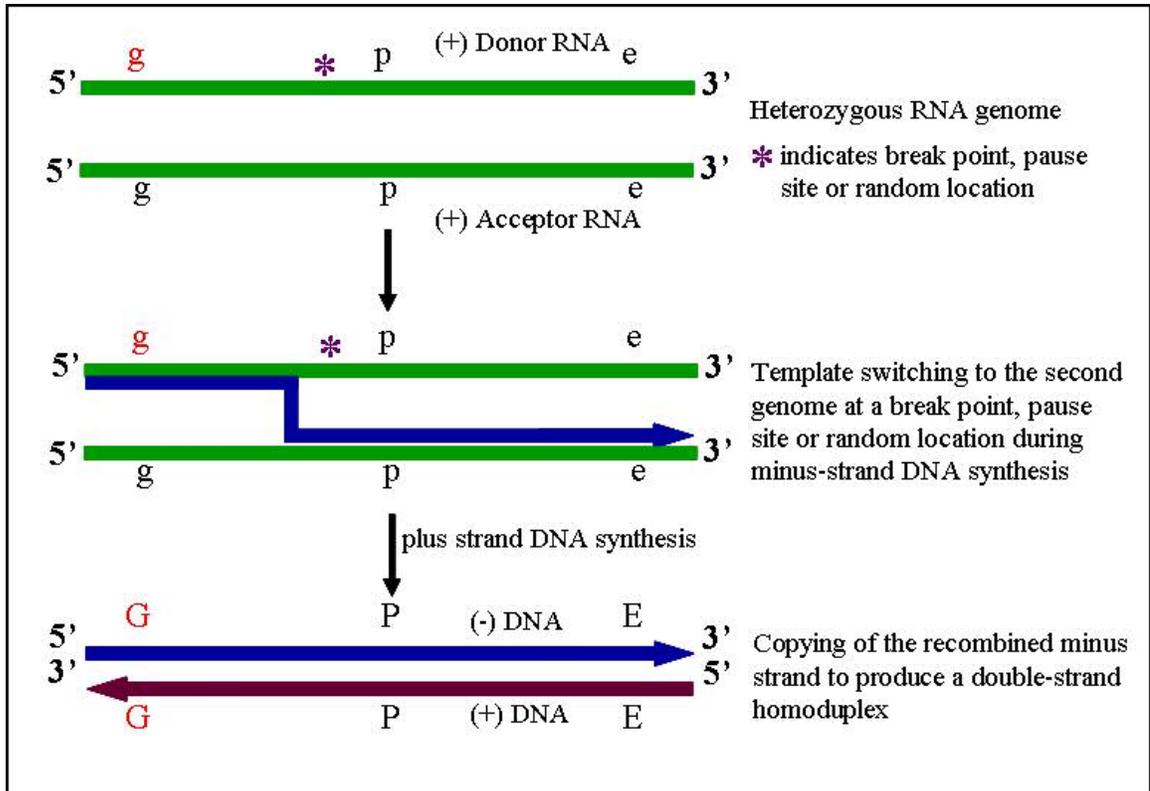


Figure 1-5: Copy-choice model

Shown here is the copy-choice model. This model proposes a mechanism for strand transfer during minus strand DNA synthesis. HIV genes *gag*, *pol* and *env* and their alleles are represented as *g*/*G*, *p*/*P* and *e*/*E* respectively. Figure adapted from Principles of Virology, Molecular Biology, Pathogenesis and Control.

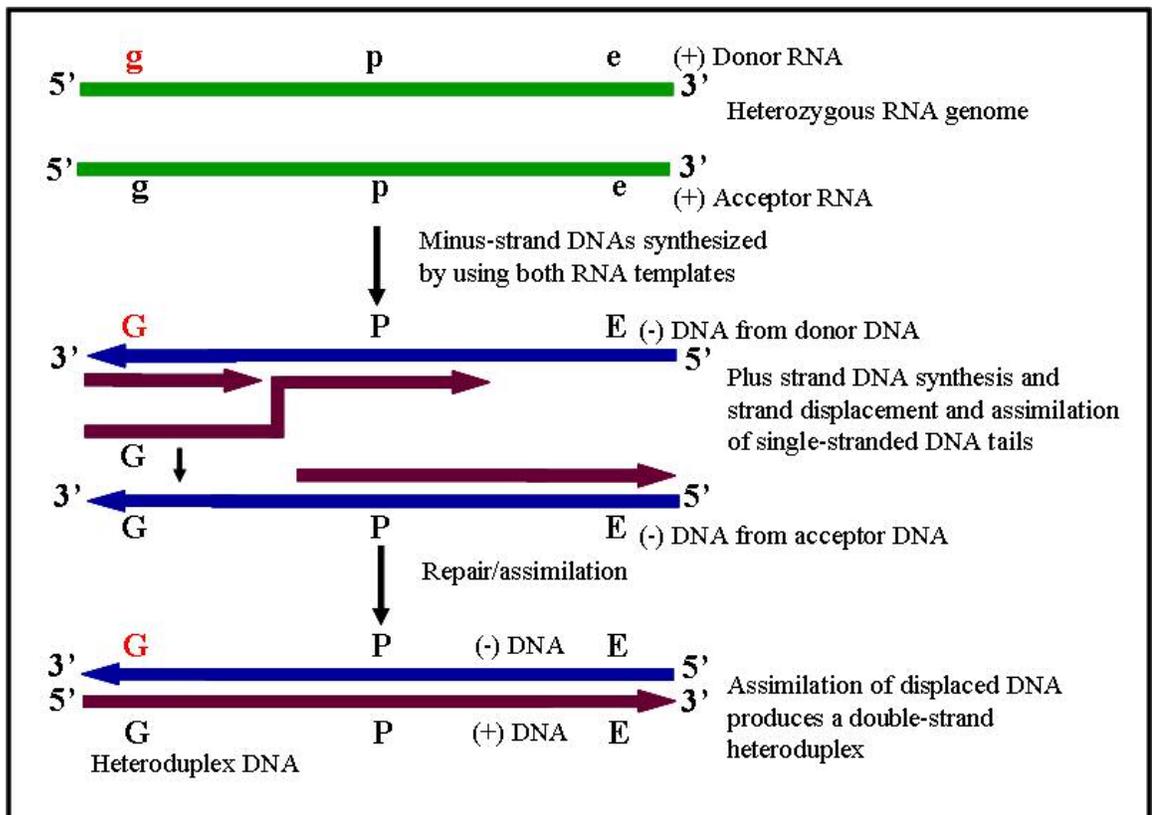


Figure 1-6: Strand displacement-assimilation model

Shown here is the strand displacement-assimilation model. This model proposes a mechanism for strand transfer during plus strand DNA synthesis. HIV genes *gag*, *pol* and *env* and their alleles are represented as *g/G*, *p/P* and *e/E* respectively. Figure adapted from Principles of Virology, Molecular Biology, Pathogenesis and Control.

1.11 Reverse Transcriptase

Reverse Transcriptase (RT) was discovered independently by David Baltimore and Howard Temin (100, 101). RT is encoded by the *pol* gene and is a multifunctional enzyme (Figure 1-7) possessing RNA dependent DNA polymerase, DNA dependent DNA polymerase and Ribonuclease H (RNase H) activity. The latter acts to degrade the RNA that is part of an RNA-DNA duplex (102). The DNA polymerase activity of RT is similar to that of most cellular DNA polymerases. It requires a primer with a 3'-OH terminus, either DNA or RNA annealed to an RNA or DNA template (103), a divalent metal cofactor (preferably Mg^{2+}) and incorporates dNTPs forming 3'-5' phosphodiester bonds with the release of pyrophosphate. RT is a relatively slow enzyme incorporating about one nucleotide per second to the growing chain *in vitro* (104) although rates *in vivo* are slightly higher. RT has modest processivity (the average number of nucleotides incorporated during a single binding event between the enzyme and primer-templates) incorporating on average about 100 bases before dissociation. RT lacks proof reading activity in that it does not possess 3'-exonuclease activity capable of excising mispaired nucleotides (105) and hence is more error prone than cellular DNA polymerases. RT has an error rate between 10^{-4} and 10^{-5} per base per replication cycle (106, 107). Therefore, RT contributes to the high mutation rate in retroviruses. Since RT plays a very important role in reverse transcription it has been an anti-HIV drug target. There are two categories of anti-RT drugs: nucleoside analogs and non-nucleoside analogs, which are used in combination with protease inhibitors to reduce viral loads in patients.

HIV-RT exists as a heterodimer consisting of two subunits of approximately 66 and 51 kDa (p66 and p51 respectively). The p66 subunit contains both the DNA polymerase and RNase H domains and the p51 subunit contains only the polymerase domain (108). The p51 subunit, which is derived from p66 by protease cleavage, is missing the carboxy-terminal region of p66 which contains the RNase H domain. The role of p51 in the function of RT is not clear. Evidence suggests that p51 subunit is not involved in the DNA polymerase and RNase H activities of RT (109). All known activities of RT lie in the p66 subunit of RT as determined from studies in which p51 subunit was inactivated by mutations (108). Postulated roles for p51 include stabilizing and protecting p66, helping to load p66 on the primer-template, and interacting with tRNA (110, 111, 112). The polymerase and the RNase H active site of RT have a catalytic core of negatively charged amino acids that interact with the metal cofactor (Mg^{2+}). Mutagenesis studies have shown that the polymerase domain has three conserved aspartate residues, Asp-185, Asp-186 and Asp-110 that are absolutely essential for polymerase function (113, 114). The RNase H active site has four conserved residues, Asp-443, Glu-478, Asp-498 and Asp-549 (115, 116). The structure of RT is similar to other DNA polymerases exhibiting an open right hand conformation containing 3 domains- the fingers, the palm and the thumb. The nucleic acid is bound to RT such that the primer terminus is in the region between the thumb and fingers and the double-stranded region is in the palm domain. The RNase H active site is about 18 base pairs removed from the polymerase site and there is a pronounced bend in the duplex between the two sites (Fig. 1-7).

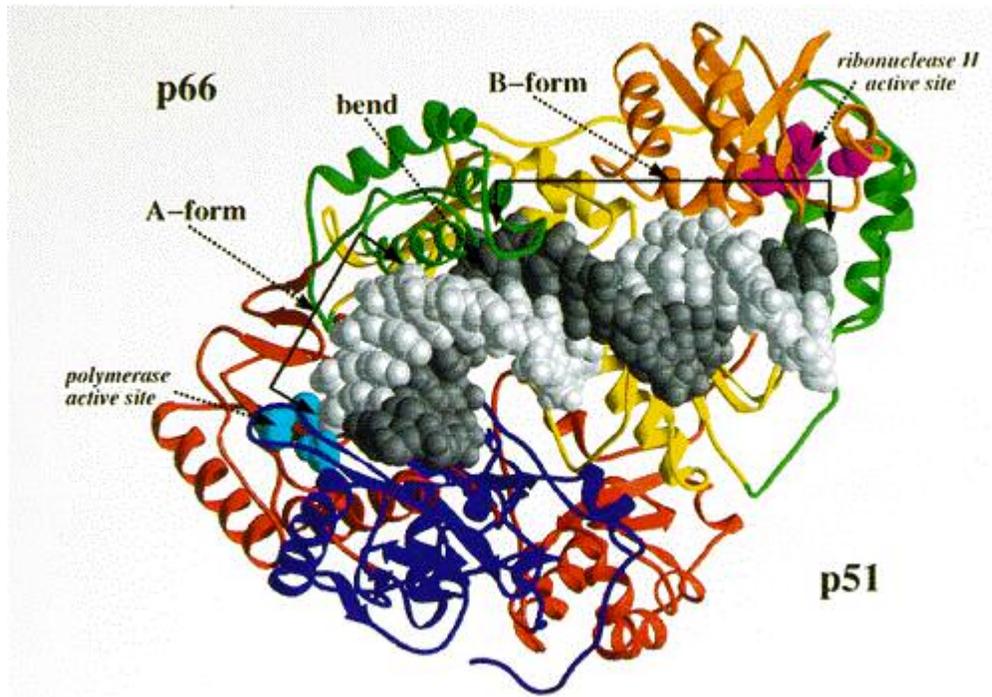


Figure 1-7: HIV Reverse Transcriptase

Shown is the ribbon diagram of HIV-1 RT bound to a space-filling model of a 19 nucleotide oligonucleotide helix substrate. The regions of A-form, B-form and bent DNAs are indicated. The various domains of RT are color-coded as: fingers-blue, palm-red, thumb-green and connection-yellow. The RNase H domain which is part of the p66 subunit is shown in orange. Figure taken from Jacoba-Molina *et al.* 1993. Proc. Natl. Acad. Sci. **90**: p 6320-6324.

1.12 Nucleocapsid protein

Nucleocapsid protein (NC) of Human Immunodeficiency Virus type 1 (HIV-1) is a small protein comprised of 55 amino acid residues (Figure 1-8). It is a highly basic and positively charged protein with a pI of 10.0 to 11.0. HIV-1, like all retroviruses encodes a *gag* gene product, which play an important role in viral assembly by recognizing and packaging two copies of viral RNA. After assembly and release of the viral particle, these *gag* gene products undergo proteolytic processing by viral protease to give several structural proteins (117). One of these proteins is the nucleocapsid protein that binds to genomic RNA (117, 118). NC proteins can bind non-specifically to single-stranded and double-stranded DNA and to single-stranded RNA *in vitro* with a preference for single-stranded molecules, it binds to RNA in the following order of affinity: retroviral RNA > mRNA > rRNA > poly(rA) (119, 120). At saturating concentrations, NC protein covers nucleic acids and offers an incomplete protection from nuclease attack (121). Initially, NC protein was purified from virions (122) but now the protein is obtained by transforming *E. coli* cells with recombinant vector and overexpression of protein (123). HIV-1 NC has two rigid zinc-binding domains or zinc fingers also called zinc “knuckles”. The two fingers are covalently linked to each other by a small flexible basic amino acid region called the linker (RAPRKKG sequence) and are flanked by flexible, basic N- or C- terminal “tails” (119, 124). Each zinc finger contains a 14- amino acid metal ion-binding motif, C-X₂-C-X₄-H-X₄-C where X denotes variable amino acids. Each zinc finger of NC coordinates one zinc ion, both *in vitro* and in virions (119, 125, 126). This tight binding of NC to zinc ions is essential to maintain stability of the zinc fingers and also ensures spatial proximity of the two fingers (127, 128). The two zinc fingers

may exhibit weak interactions with one another (129). NC is susceptible to anti-viral agents that have the ability to eject zinc, which interferes with the normal functions of NC (130-132). The two zinc fingers have similar structures (133) although the amino acid sequences surrounding the CCHC motifs are different (124). The biological activities of the two fingers are not equivalent (134, 135) and the presence of both fingers is crucial for the production of replication-competent virus (136). Also, the positions of the zinc fingers cannot be exchanged (134). NC protein of HIV has highly conserved hydrophobic residues at positions 13, 16, 24 and 25, a basic residue at position 26 and glycine residues at positions 19 and 22 (122, 137). NMR studies are used to determine the three-dimensional structure of NC protein. It has been shown that NC binds to several nucleic acid targets during the viral life cycle. The hydrophobic residues of NC located in the second position of each zinc finger exhibit strong stacking interactions with nucleic acid bases (138). This stacking was found to be most efficient with G bases, especially when preceded by T (139, 140). Binding of NC to stem-loop sequences (SL1, SL2, SL3, SL4) that constitute the ψ genomic packaging signal were studied. Close interactions were observed between Phe16 and Trp37 and purine residues in single stranded regions of SL2 and SL3 RNA hairpins in ψ genomic packaging signal of HIV-1 (141, 142). Binding of NC to SL1 and SL4 is weak. In addition, DNA analogs of the RNA stem-loops bound to NC less efficiently than the corresponding RNA (140). NMR studies also showed that basic residues of NC help in formation of intramolecular salt bridges that stabilize folding of NC zinc fingers and also participate in forming electrostatic interactions with the RNA (141, 142). Computational studies were done to determine NC's nucleic acid binding properties and the ability of NC's zinc co-ordinating Cys residues of finger 2 were found

to be more susceptible to electrophilic attack with Cys49 of finger 2 being the most labile to electrophilic attack which is in agreement with experimental observations (143).

NC exhibits several important functions in the life cycle of HIV. These functions can be attributed to its 'nucleic acid chaperone' activity. This activity allows NC to catalyze the rearrangement of nucleic acids into a more thermodynamically favorable conformation (124, 144). For NC, the chaperone activity has two components: helix-destabilization and condensation/aggregation. Studies with mutant proteins showed that the basic regions of NC are necessary for NC chaperone activity (145, 146). Work has shown the helix-destabilization also requires the zinc fingers with finger 1 playing the major role (134, 135). The condensation/aggregation activity is probably due mostly to the basic nature of NC as cations like poly-lysine and spermidine also show this effect on nucleic acids (147). This activity is responsible for NC's ability to increase the rate of annealing, even for complements without secondary structure (135). NC and other chaperones appear to function by causing transient unpairing of bases. These bases are now free and available for pairing with other bases to attain the most favorable conformation thereby permitting the bases to escape from less energetically favorable conformations. NC's chaperone activity aids in a variety of functions like nucleic acid unwinding, nucleic acid annealing and strand exchange. NC can thereby promote unwinding of structured RNAs like tRNA and the TAR stem-loop of the viral RNA and annealing of complementary strands as well as facilitate transfer of nucleic acid from a thermodynamically less stable hybrid to one that is more favorable (119, 144). In this

regard, NC behaves like the single strand DNA binding (SSB) proteins of *E. coli* and the RNA binding proteins like heterogeneous nuclear ribonucleoprotein A1 (148). The participation of NC has been implicated in almost every process of the viral life cycle. *In vitro* studies have shown that NC has several important roles in reverse transcription. Reverse transcription begins when the 18 nucleotides near the 3' end of host cell derived tRNA^{Lys3} anneals to a complementary region on the genomic RNA of HIV called the primer binding site (PBS). NC helps in the unwinding of the 3' end of tRNA^{Lys3} (149, 150) and also helps its annealing to the PBS (151-153). During reverse transcription, there are two obligatory strand transfers that occur at the termini of the genomic RNA without which reverse transcription cannot proceed to completion (83, 84, 86). NC protein enhances these obligatory transfers (154-157). In the first strand transfer, the -ssDNA that is synthesized by 3' extension of the tRNA primer bound near the 5' end of the genome is translocated to the complementary R region at the 3' end of the genome for continuation of synthesis. It has been found in *in vitro* assays that the first strand transfer is about 25 times more efficient in the presence of high concentrations of NC (124). The second strand transfer involves the +ssDNA that is generated by extension of the polypurine tract up to the modified base on tRNA. For synthesis to continue, this tRNA must be removed so that the complementary PBS sequences on the minus strand and plus strand DNA can anneal. NC has been shown to enhance both tRNA removal and aid in annealing of the complementary PBS sequences (158).

In addition to the two obligatory transfers, the growing DNA molecule can transfer to another template at any internal position along the genome. This is called an

internal strand transfer and is the mechanism by which recombination occurs. The copy-choice model described earlier in this chapter explains these internal strand transfers. It has been shown that secondary structures or sequences along the genome make it difficult for RT to read through. These regions are referred to as 'pause sites' as they cause RT to pause during reverse transcription. At these sites, synthesis can continue if the stalled DNA transfers to the second RNA of the virion. NC is found to reduce pausing at some but not all pause sites (159, 160) and also facilitates transfer of the nascent DNA from these sites (161, 162). This activity of NC is due to its helix-destabilizing and nucleic acid annealing properties (155, 160, 163-65). NC is also known to have a modest positive effect on the processivity of reverse transcriptase during DNA synthesis, probably by making it easier for RT to traverse secondary structures (160, 166, 167). Another important aspect of NC's chaperone activity is its ability to aggregate nucleic acids thereby facilitating attraction between nucleic acid strands (121, 168-171). Complementary sequences can thereby search for each other within such aggregates. Work done in this thesis shows how aggregates formed by NC play an important role in the synthesis of long DNA products from genomic RNA of HIV *in vitro*.

After reverse transcription, NC has been shown to play a role in the integration of proviral DNA into the host chromosome by stimulating IN activity (121, 172-174). Also, NC sequences in the Gag precursor specifically interacts with an intact packaging signal on the viral genome (141) and participates in recognition and packaging of viral RNA into the virion (134, 175-177). NC has also been implicated in the dimerization of the two viral RNAs and the maturation of this genomic RNA dimer (178, 179).

The important functions carried out by NC during replication and infection make it an ideal target for drug therapy (180-182) and vaccine development (§3). Work done in this thesis shows how NC plays an important role in the synthesis of long DNA products from HIV genome regions *in vitro*.

1.13 Goals: Improving *in vitro* reverse transcription reactions and understanding how RT produces long DNAs in the cell

Reconstituted *in vitro* reverse transcriptase reactions have greatly contributed to the current understanding of the reverse transcription process. However, before this work the systems did not faithfully mimic reverse transcription in the cell with respect to the synthesis of long DNA products (184). In the cell synthesis of an approximately 10,000 nucleotide minus strand from the RNA genome must occur. *In vitro* reactions were only capable of producing RNAs of a few thousand nucleotides with the vast majority of products being less than 1,000 in reactions with templates of several thousand bases. There are several possible reasons for this including cellular or virion factors that may be missing *in vitro*, or a structural framework provided by virion components in the replication complex. For example, the complex could provide a tightly packed and condensed environment that helps RT overcome its low processivity. The goal of the current thesis was to generate *in vitro* reaction conditions that produced DNA products that are several thousand nucleotides long. This *in vitro* system could be used to screen inhibitors of NC, RT and recombination. Such a system could also be useful for synthesis of long cDNAs from RNA.

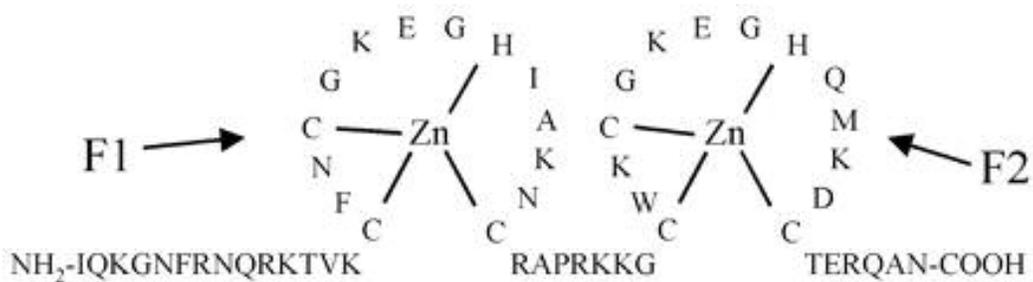
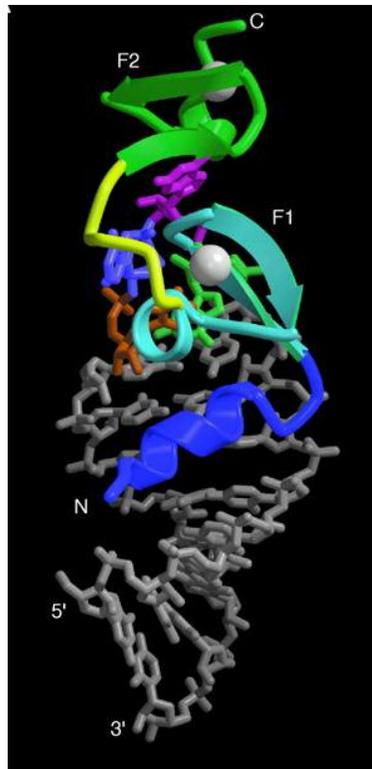


Figure 1-8: HIV Nucleocapsid protein

Shown above is a ribbon diagram of HIV-1 NC-SL3 Ψ-RNA complex. The color codes: 3₁₀ helix-purple, first zinc finger-blue, second zinc finger-green, zinc atoms-white spheres, RNA-gray except the guanosine residues which are colored (Figure taken from Ref 142). Shown below is the 55 amino acid sequence of HIV-NC protein (Figure adapted from Ref 124). The amino and carboxyl terminals are shown. Each NC molecule has two zinc fingers F1 and F2, each of which has the C-X₂-C-X₄-H-X₄-C motif, where C denotes the variable amino acids. The two fingers differ from each other by five amino acid residues.

Chapter 2 Formulation of reaction conditions *in vitro* for the synthesis of long reverse transcription products from genomic RNA segments of Human Immunodeficiency Virus (HIV)

2.1 Introduction

Human Immunodeficiency virus (HIV-1), like all retroviruses, undergoes reverse transcription during its replication cycle. This process involves copying of its single stranded RNA genome into double-stranded DNA that later gets integrated into the host chromosome. Although some viral particles may initiate reverse transcription before entering cells (185-189), the process generally begins when the viral particle enters the cytoplasm of the host cell and occurs exclusively in the host cytoplasm. HIV cores are very fragile and are disrupted shortly after virus-cell fusion and entry into the host cell (190). The disassembly of the HIV core is crucial for the initiation of reverse transcription (191). The viral RNA and associated proteins are released into the cytoplasm and can interact with the cytoskeleton. They were found to specifically localize with actin microfilament components and disruption of this microfilament cytoskeleton inhibited reverse transcription in cells (192). Although HIV cores are dissociated shortly upon viral entry, reverse transcription is known to occur within a confined environment in capsid-like structures permeable to dNTPs and derived from the virion core (193, 194). Such a structure is referred to as reverse transcription complex (RTC). The RTC perhaps provides an optimal environment by preventing loss of factors required for reverse transcription and also offers protection from cellular nucleases.

Reverse transcription intermediates and products identical to those made in infected cells can be synthesized in the purified virions. This requires application of mild detergents to permeabilize the envelope and addition of deoxyribonucleoside triphosphates and a divalent cation as the metal cofactor (referred to as endogenous reverse transcription or ERT) (195-198). This process can even occur in the absence of added detergent (referred to as natural endogenous reverse transcription or NERT) (199-201). In this case, the amphipathic domains of C-terminus of gp41 make the envelope naturally permeable to dNTPs (201). However, the yield of completed products in ERT and NERT reactions is extremely low. It can be concluded that the virion environment is not sufficient to allow efficient replication or that certain cellular components or structural alterations are essential for efficient replication.

Reverse transcription can also be carried out *in vitro* in totally reconstituted reactions including only a primed RNA template and purified reverse transcriptase (RT). Several effects of the viral nucleocapsid protein (NC) have been demonstrated in these reactions including: enhancing binding of the host tRNA primer to the primer binding site on the viral RNA (as part of the gag precursor protein) (150, 202, 203), increasing the processivity of RT (166, 204) promoting dimerization between the two genomes in the viral capsid (205, 206) and stimulating strong-stop minus and plus strand transfer and viral recombination in general (154-162). In fact, all the reactions and steps required to produce complete double-stranded DNA appear to reside in RT and viral NC enhances many of these reactions. It has also been suggested that other viral proteins (Tat, for example) may play a role, but are probably not major constituents of the replication

complex (207). Also, both RT and NC are highly stable *in vitro*. Therefore, given enough time, one would predict the production of completely processed double-stranded DNA from genomic RNA in *in vitro* systems. However, this is not observed and even fully synthesized single stranded DNAs (minus strand) are not generated from genome length RNA. Most products are only a few hundred and at most a few thousand nucleotides in length when templates of several thousand bases are used (184), indicating that these reactions are even less efficient than NERT or ERT reactions. This could be because the capsid creates an extremely concentrated and specific environment that cannot be completely reproduced *in vitro*. The concentrated environment within the capsid may promote rebinding of nucleic acids and/or reverse transcriptase by keeping molecules in close proximity, and may constrain the template and help preserve the replication intermediate. In the more dilute test tube environment, there could be a greater chance of dissociation of the replication intermediates and more difficulty of rebinding after dissociation.

In the presence of specific amounts of NC and high RT concentrations an increase in reverse transcription efficiency in *in vitro* reactions can be observed. Results showed that at greater than or equal to 50% saturation of NC binding sites (one NC coats approximately 7 nucleotides(120, 123, 208), there is up to a 90% decrease in total DNA synthesis, as measured by incorporation of dNTPs (209). However, the cDNA products made are almost exclusively full-length, although in this case an RNA template of only 874 bases was used.

As part of an effort to produce an *in vitro* system that more closely mimics cellular replication, in this report conditions that produced DNA products up to 4 kb from genomic RNA of HIV at relatively high efficiency are described. These reactions included high concentrations of RT and enough NC to completely coat all the nucleic acids in the reaction. In addition to enhancing our understanding of the replication process, an *in vitro* system that mimics cellular replication could potentially be used to screen reverse transcription inhibitors.

2.2 Materials

Plasmid pBKBH10S was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH, from Dr. John Rossi. This plasmid contains an 8.9 kb SstI fragment (nt 222-9154 of the RNA genome) from HIV-1 BH10 inserted into the SstI site. The fragment has all HIV-1 gene coding regions but does not contain the HIV-1 LTR (210). Plasmid pNL4-3 was also obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: pNL4-3 from Dr. Malcolm Martin. This plasmid contains a full-length, replication and infection competent chimeric DNA i.e. the 5' SmaI-EcoRI fragment of proviral NY5 (5' SmaI in flanking sequences to 3' EcoRI) and the 3' fragment of proviral LAV (5' EcoRI to 3' NruI in flanking sequences) that was blunt-end cloned into pUC18 at the PvuII site after removal of polylinker sites. Plasmid pBR322 with hepatitis delta virus (HDV) ribozyme and 2 tandem T7 terminators (pBR322dR) was a gift from Dr. Siba Samal, VA-MD Regional College of Veterinary Medicine. PCR primers and primers used to prime templates in reverse transcription assays were obtained from Integrated DNA Technologies, Inc. The HIV-RT clone was a generous gift from Dr. Samuel H. Wilson (National Institute of Environmental Health Sciences, Research Triangle Park, NC). HIV-RT was purified according to the protocol described (211). The protein was purified to homogeneity and the purity of the protein was evaluated using Coomassie Blue staining of 10 % SDS-PAGE gels (212). The subunits p51 and p66 of RT were in a 1:1 ratio. Aliquots of HIV RT were stored frozen at -80°C and fresh aliquots were used for each experiment. The HIV NC clone was a generous gift from Dr. Charles McHenry (University of Colorado). NC was purified to apparent homogeneity (as judged from

Coomassie Blue staining of 17.5% SDS-PAGE gels (212)) according to the protocol described (123). Quantification was by absorbance at 280 nm using a molar extinction coefficient of $8350 \text{ cm}^{-1} \text{ M}^{-1}$ (123). Aliquots of NC were stored frozen at $-80 \text{ }^{\circ}\text{C}$, and fresh aliquots were used for each experiment. *Taq* polymerase was from Eppendorf. T7 RNA polymerase, DNase I-RNase-free and RNase-DNase-free were from Roche Diagnostics. RNase inhibitor was from Promega. T4 polynucleotide kinase and Restriction enzymes EcoRI, HincII, Asc I, Age I and Rsr II were obtained from New England Biolabs. Proteinase K was obtained from Eastman Kodak Co. Radiolabeled compounds were obtained from Amersham. Sephadex G-25 spin columns were from Amika Corp. RNA cleanup kit was from Qiagen. All other chemicals were from Sigma or Fisher Scientific.

2.3 Methods

PCR Amplification of DNA substrates for cloning into pBR322 plasmid with HDV ribozyme and 2 tandem T7 terminators - Two PCR primers, 5'-GAT**CGGCGC**GCTAATACGACTCACTATAGGGGGGTCTCTCTGGTTAGACCA GATCTG -3' and 5'- AAATTTGATATGTCCATTGGCCTTG -3' were designed to amplify DNA from position 1 to position 3573 of the HIV insert on pNL4-3. The sequence indicated in bold is the recognition site for restriction enzyme AscI. Also, two PCR primers, 5'- GCACTAACAGAAGTAGTACCACTAAC-3' and 5'-AGCT**CGGACCG**CGAGGAGGTGGAGATGCCATGCCGACCCTTAATCCTCATCC TGTCTACTTGCCAC-3' were designed to amplify DNA from position 3411 to 5096 of the HIV insert on pNL4-3. The sequence indicated in bold is the recognition site for restriction enzyme RsrII. and used for cloning as described below.

Preparation of plasmid pBR322 with HDV ribozyme and T7 terminators (pBR322dR) containing the HIV sequences- The construct developed to make genomic RNA segments of HIV was prepared by inserting fragments of the HIV sequence derived from pNL4-3 which contains a cDNA copy of HIV genome cloned into pUC18 as described above. The PCR products obtained using the first set of primers described above were digested with AscI and AgeI. The PCR products obtained using the second set of primers described above were digested with AgeI and RsrII. The construct was designed by first inserting a 3485 base pair AscI-AgeI PCR fragment (obtained as described above) into the AscI and AgeI site on pBR322dR. The second 1611 base pair AgeI-RsrII PCR fragment (obtained as described above) was then inserted into the AgeI

and RsrII sites on pBR322dR to get a 5096 base pair cDNA copy of HIV genome in pBR322dR.

Preparation of RNA substrates- RNAs of approximately 1.9 and 4 kb were made by first digesting pBKBH10S with restriction enzymes HincII and EcoRI, respectively. RNA of approximately 5.1 kb was made by first digesting pBR322dR with the HIV insert (described above) with AseI. The digests were then extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with ethanol. Run-off transcription (performed according to the enzyme manufacturer's protocol) was then conducted using 5 µg of the digest plasmid and T7 RNA polymerase enzyme to generate 1.9, 4 and 5.1 kb RNAs. The transcription reactions were treated with 2 µl of 10 units/µl of DNase I-RNase-free enzyme for 15 min to digest away the template DNA. The RNA was purified using the Qiagen RNA cleanup kit. The amount of recovered RNA was determined spectrophotometrically from optical density. The integrity of the RNA was checked on a 1% agarose gel. Typically, two bands corresponding to 800nts and 1200 nts of dsDNA were seen with the 1.9 kb RNA

RNA-DNA Hybridization- DNA primers that bound specifically to the RNA templates: 5'-CTGAAGCTCTCTTCTGGTGG-3' to the 1.9 kb template and 5'-GCTTGATTCCCGCCACCAA-3' to the 4 kb template were ³²P-labeled at the 5'-end with T4 polynucleotide kinase according to the manufacturer's protocol. The 1.5 kb RNA template was hybridized to its complementary labeled primer by mixing primer: transcript at a ~ 1:1 ratio and the 4 kb RNA template was hybridized to its

complementary labeled primer by mixing primer: transcript at a ~ 4:1 ratio in 50 mM Tris-HCl (pH 8.0), 1mM dithiothreitol and 80 mM KCl. The mixture was heated to 70°C for 5 min and then slowly cooled to room temperature.

Reverse transcription reactions with or without NC- RNA template-DNA primer hybrids (4 nM final concentration of RNA) were pre-incubated for 5 min along with additional template RNA (12 nM) in the presence or absence of NC (4 μ M) in 21 μ l of buffer (see below) at 37°C. The reactions were initiated by addition of 4 μ l of HIV-RT (80 nM final in reactions). The following reagents at the indicated final concentrations were also included in the reaction mixtures: 50 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 80 mM KCl, 6 mM MgCl₂, 100 μ M dNTPs, 5 mM AMP (pH 7.0), 25 μ M ZnCl₂ and 0.2 units/ μ l RNase inhibitor. Reactions were allowed to incubate for 75 min. The reactions were stopped by adding 2 μ l of a solution containing 250 mM EDTA (pH 8.0) and 5 ng of RNase-DNase-free enzyme and allowed to digest for 20 min at 37°C. Nine μ l of proteinase K at 2 mg/ml in 1.25 % SDS, 15 mM EDTA (pH 8.0) and 10 mM Tris (pH 8.0) was then added to the above mixture, which was placed at 65°C for 1 hour. Finally 7 μ l of 6X alkaline dye (300 mM NaOH, 6 mM EDTA, 15% glycerol, 0.15% bromophenol blue) was added to the mixture and the samples were resolved on 1% alkaline agarose gel containing 50 mM NaOH and 1 mM EDTA (pH=8). Similar reactions were conducted for enzyme titration experiment and for reactions with excess template (which were carried out with increasing amounts (4, 8, 12 and 32 nM) of template (1.9 kb)) Extended DNA products were observed using a Bio-Rad Molecular Imager FX.

Time course reaction- The 1.6 kb RNA template was hybridized to DNA primer as described above (4 nM final concentration of RNA in reaction) and was pre-incubated for 5 min along with acceptor RNA (12 nM) in the presence of NC (4 μ M) in 21 μ l of buffer (as in reverse transcription experiment) at 37°C. The reactions were initiated by addition of 4 μ l of HIV-RT (80 nM final in reactions). The reactions were stopped as described above at time points- 2, 5, 10, 15, 30, 45, 60 and 75 mins and samples were resolved on 1% alkaline agarose gel containing 50 mM NaOH and 1 mM EDTA (pH=8). Extended DNA products were observed using a Bio-Rad Molecular Imager FX.

Gel electrophoresis- One percent alkaline agarose gels containing 50 mM NaOH and 1 mM EDTA (pH=8), 1 % native agarose gels in Tris-Borate-EDTA buffer were prepared and subjected to electrophoresis as described (212).

2.4 Results

Optimization of reaction conditions for the synthesis of long DNA products in vitro- The general approach used for the reverse transcription assay is depicted in Figure 2-1. RNA templates for the reverse transcription assay were derived from pBKBH10S plasmid by run-off transcription. These templates were then hybridized near their 3' end to 5' radio labeled primers and reaction was initiated by addition of RT enzyme.

Synthesis of the 1.5 kb DNA product from HIV genomic RNA in vitro- Reaction conditions were optimized first for the synthesis of a 1538 nucleotide DNA product *in vitro*. This was done by performing the reverse transcription assay with different concentrations of NC (Figure 2-2) and also with different concentrations of HIV-RT (Figure 2-3). In the reverse transcription assay with different NC concentrations, some fully extended products were observed with 2 μ M NC and most products were full-length when 4, 6, or 8 μ M NC was used. Reactions with one μ M NC resembled those without NC. It can be noted here that in the absence of NC, there is primer extension resulting in synthesis of small DNAs and no full-length products. In the presence of NC, there is a decrease in total synthesis and most products formed are full length (Figure 2-2). With 2, 4, 6, and 8 μ M NC, primer extension was 80, 62, 40, 16%, respectively, of reactions without NC. The level of extension with NC varied to some extent depending on the RNA preparation used and other factors and was typically between 35-70% with 4 μ M NC. The decrease in primer extension should be accompanied by a corresponding increase in unextended primer, but this is not typically observed because a large portion of the small primers diffuse out of the alkaline agarose gel during electrophoresis and

processing. The unextended primers were evident on denaturing polyacrylamide gels although large products cannot be resolved on these gels (data not shown). The reactions in this assay included 16 nM of the 1.9 kb RNA template and 4 nM 20 nucleotide primer or about 26 μM total nucleotide. Assuming one NC molecule can coat about 7 nucleotides (120, 123, 208) and most molecules are bound to nucleic acid under the conditions, approximately 3.7 μM NC would be required for complete coating. The appearance of mostly long products at 4 μM NC and essentially no long products at 1 μM were consistent with complete or nearly complete coating being required for their production. At very high NC concentrations (8 μM), there is a decrease in synthesis of full-length DNA products due to inhibition by NC of synthesis at very high concentrations.

In the reverse transcription assay with different concentrations of RT enzyme, there is an increase in full-length DNA formed with increasing enzyme concentration. At least 80 nM RT was required for optimum production of full-length products. When RT concentrations are further increased (160 nM) as shown, full length DNA synthesis did not improve indicating that 80 nM RT enzyme was the optimal concentration for synthesis of full-length DNA products under the conditions used in this assay. Hence 80 nM RT was used in the experiments.

Time course assay- A time course reverse transcription assay was performed to determine the time required for the generation of full-length DNA products. This was done by performing the reaction with 4 μM NC and 80 nM RT (concentrations of NC and

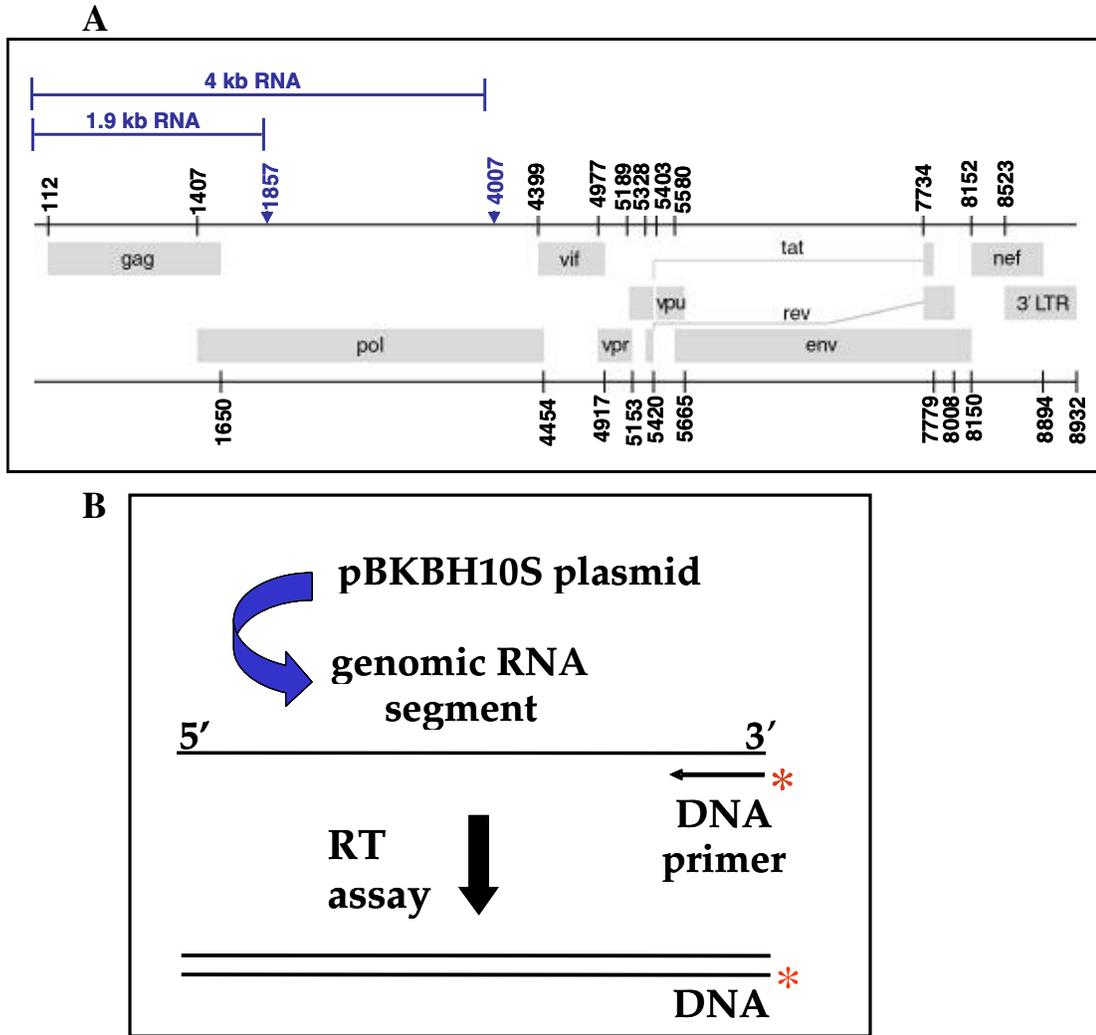


Figure 2-1: A) Schematic representation of HIV-1 BH10 genome (8932 nucleotides- from nucleotide 222-9194 of HIV-1 HXB2) in pBlueScript II KS+. The 1.9 kb RNA is 1857 nt (from 222-1759 on HIV-1 HXB2) and the 4kb RNA is 4007 nt (from 222-4228 on HIV-1 HXB2) B) Schematic representation of the reverse transcription assay. Genomic RNA used as template for the reverse transcription assay was derived from pBKBH10S plasmid by run-off transcription. The RNA was hybridized to a radiolabeled DNA primer and the assay carried out by adding RT. The DNA products obtained were run on a 1% alkaline agarose gel. The template lengths were 1857 (1.9 kb) and 4007 (4 kb) nucleotides and fully extended DNA products were 1538 and 4007 nucleotides for the short and long templates, respectively.

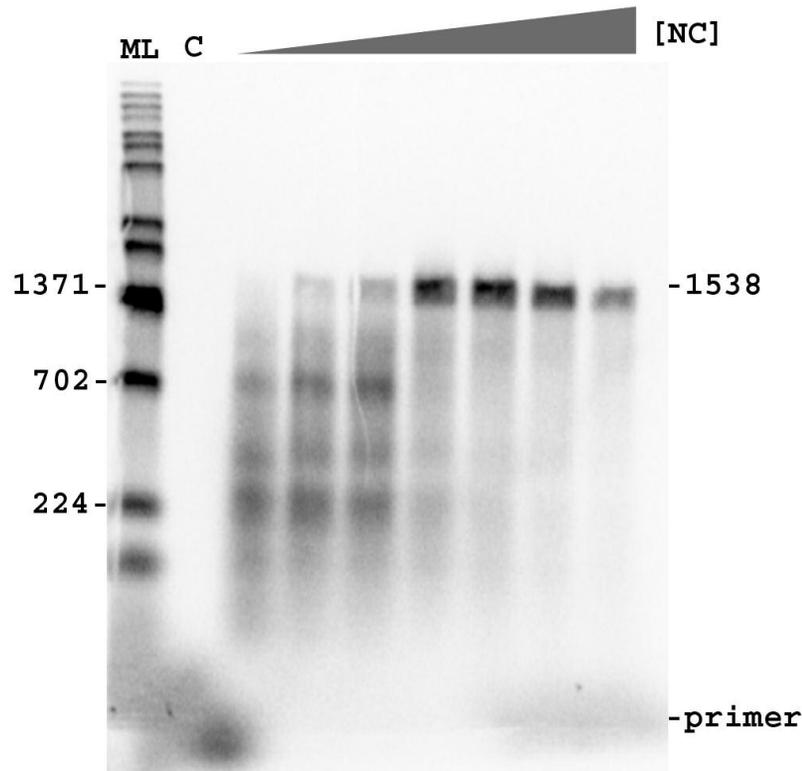


Figure 2-2: Reverse transcription assay with different concentrations of wild type NC using 1.9 kb RNA as template.

Shown is an autoradiogram of an assay with 1.9 kb RNA using different NC concentrations (0.5, 1, 2, 4, 6 and 8 μM) or in the absence of NC (0 μM) performed under conditions as indicated under “Methods”. Positions of size marker (in nucleotides) are shown on the left and full-length products are indicated on the right (“1538”) along with the primer location.

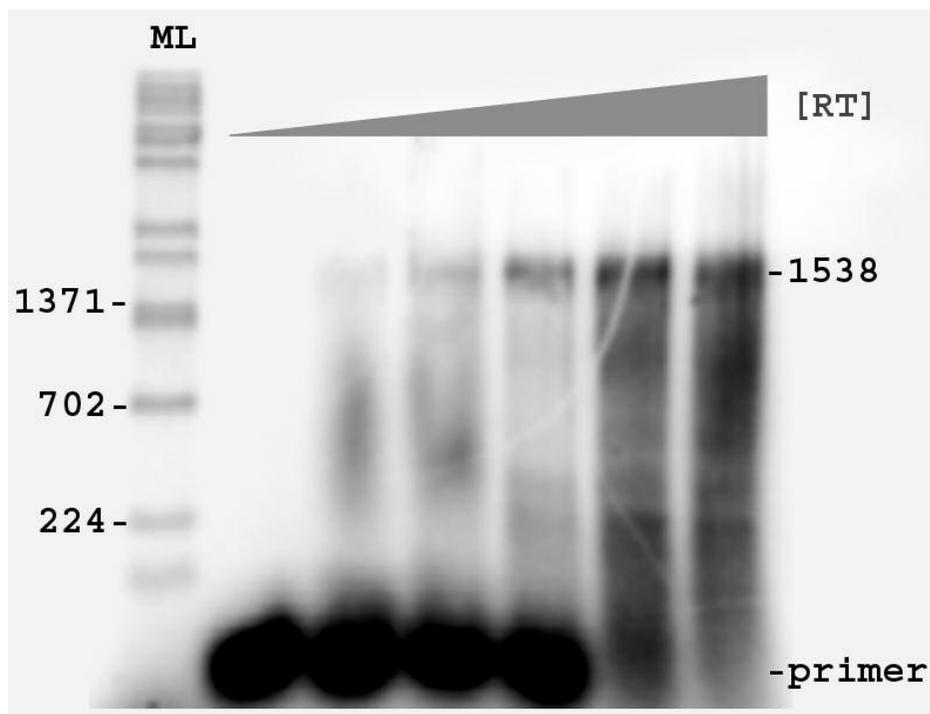


Figure 2-3: Reverse transcription assay with different concentrations of wild type RT using 1.9 kb RNA as template.

Shown is an autoradiogram of an assay with 1.9 kb RNA using different RT concentrations (0, 10, 20, 40, 80, 160 nM) in the presence of NC (4 μ M) performed under conditions as indicated under “Methods”. Positions of size marker (in nucleotides) are shown on the left and full-length products are indicated on the right (“1538”) along with the primer location.

RT that were found to be optimum from NC and RT titrations respectively) and stopping the entire reactions at the end of 2, 5, 10, 15, 30, 45, 60 and 75 min (Figure 2-4). The full-length products are observed only after 45 minutes of the start of reaction and increased up until 60 min with no further increase at 75 min indicating that it takes at least 45 min for the full-length DNA products to be synthesized. Given the approximately 1500 nucleotide size of the template this implies a maximum synthesis rate of about 33 nucleotides per minute in the reactions. This rate is somewhat lower than the 150-180 nucleotide/min estimate for cellular synthesis during infection (213).

Increase in synthesis of full-length DNA is observed with an increase in template concentration- Since the amount of primer extension was lowered in the presence of NC, experiments were conducted to determine if adding different amounts of template could change the level of extension. Varying the template would change the number of primer binding sites. The amount of fully extended products and total primer extension were increased in the presence of excess template (Figure 2-5). At 1:1 template: primer (4 nM each) only a small amount of extended primer was observed (34% of –NC reaction). The amount of primer extension increased with 42, 52, 70, and 83% of –NC reactions observed with 2:1, 3:1, 4:1 and 8:1 template:primer, respectively. The experiment shows that the apparent inhibition of primer extension in the presence of NC can be mostly overcome by adding extra primer binding sites.

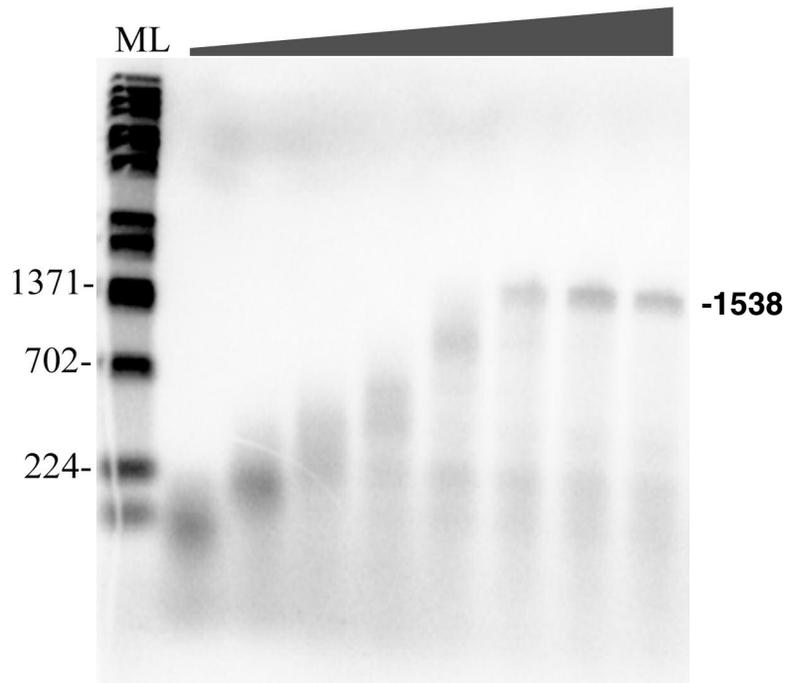


Figure 2-4: An autoradiogram of a time course experiment.

The reverse transcription reactions were carried out in presence of 4 μ M NC for time points 2, 5, 10, 15, 30, 45, 60 and 75 mins as shown from left to right. The positions of full length DNA products from the donor 1538) is indicated. ML denotes lane with molecular marker (in nucleotides).

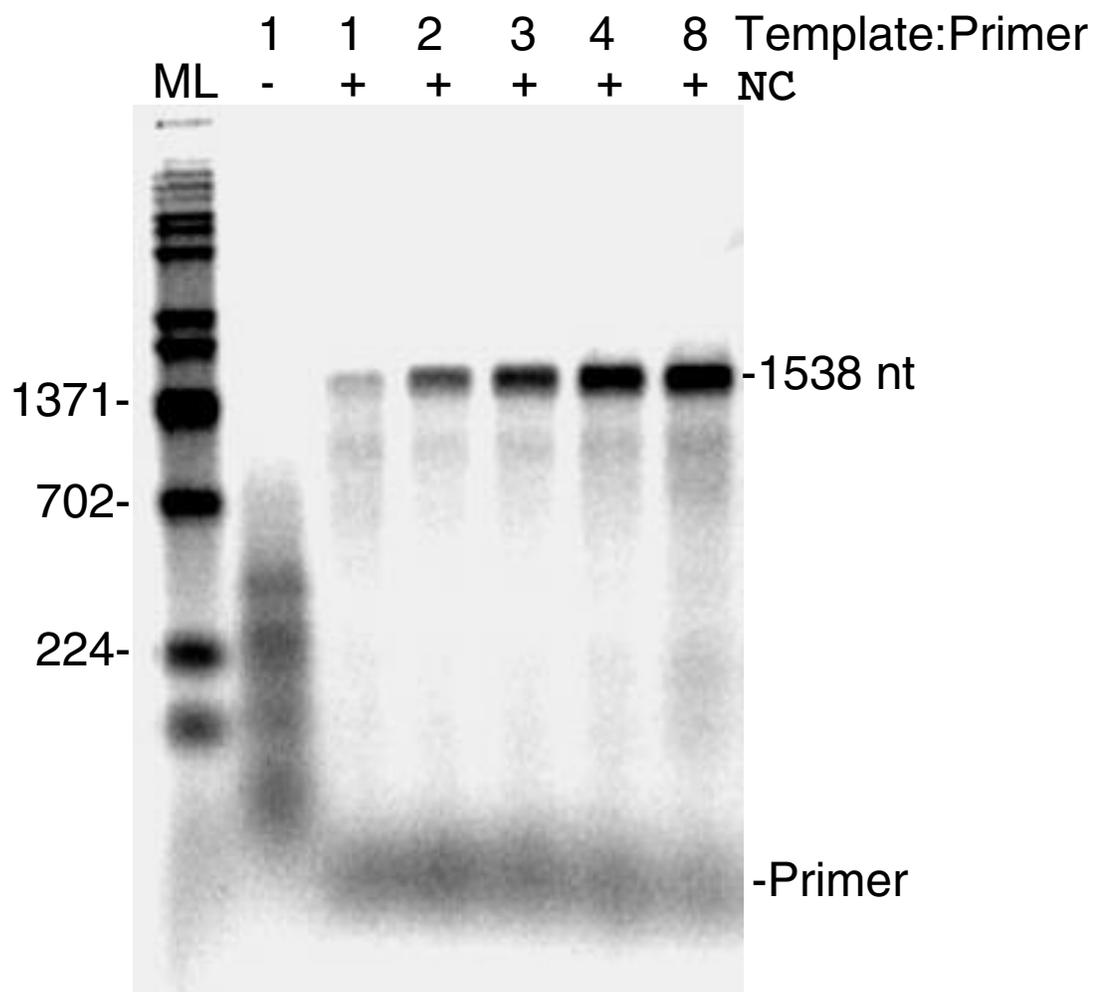


Figure 2-5: Reverse transcription assay with increasing concentrations of template.

Shown is an autoradiogram of an assay using 1.9 kb RNA as template with no excess template (1:1 template: primer) and with increasing concentrations of template (2, 3, 4, and 8 template: primer). Lanes with – indicate reactions without NC and those with + indicate reactions with NC (4 μ M). ML denotes lane with molecular marker (in nucleotides).

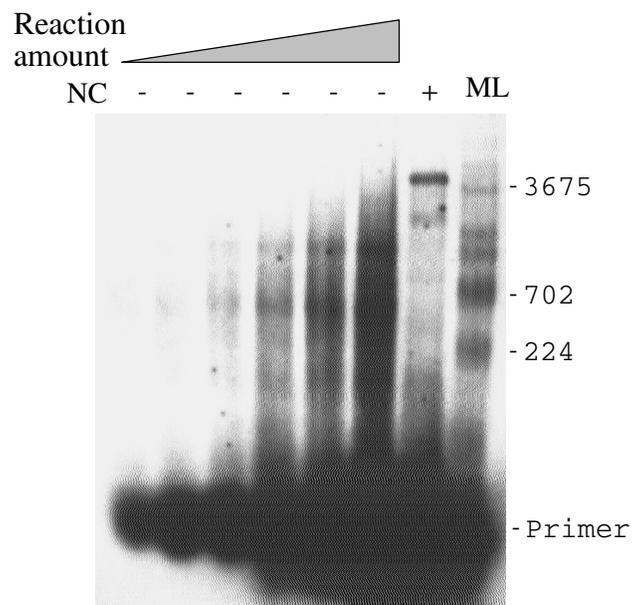


Figure 2-6: Reverse transcription assay using 4 kb genomic RNA segment as template.

Reactions are performed with (+) or without (-) NC (4 μ M) as indicated. Increasing amounts of reaction material were loaded in lanes with reactions without NC. The amount of material in the last lane without NC is equal to the amount in the lane with NC. ML denotes lane with molecular marker (in nucleotides).

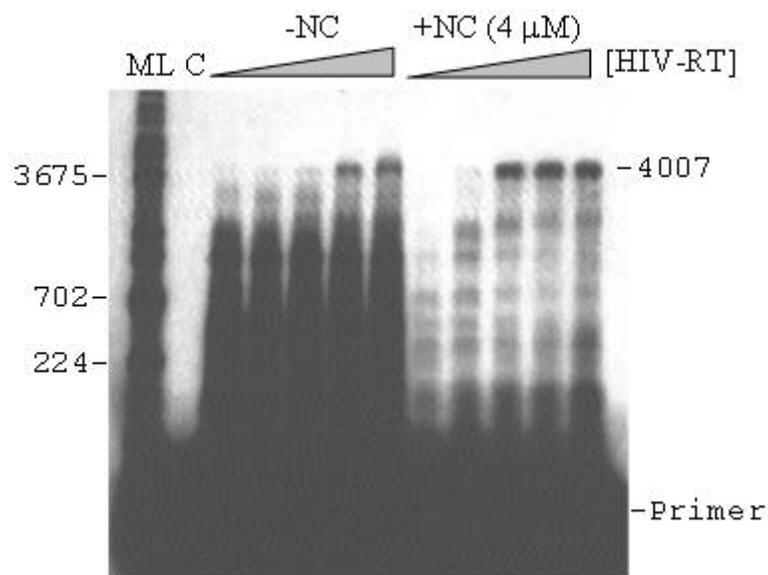


Figure 2-7: RT titration assay using 4 kb genomic RNA segment as template.

Reactions with or without NC (as indicated) are shown with increasing concentrations of HIV-RT. The amounts used were (from left to right) 20, 40, 80, 160 and 320 nM. The position of full length DNA products is indicated (“4007”). ML denotes lane with molecular marker (in nucleotides).

Synthesis of the 4 kb DNA product from HIV genomic RNA in vitro- Reverse transcription was then performed using a longer RNA on which fully extended primers produced products of approximately 4 kb. Various experiments were performed to characterize these reactions and to optimize reaction conditions. The reactions required high amounts of RT and NC. With the longer template also, enough NC to coat all the RNA in the reactions was required based on one molecule of NC coating 7 bases. The amount of NC required to produce optimal results varied to some extent from one RNA preparation to the next and the “window” was relatively small with inhibition observed at high concentrations. Reactions worked best using synthesis conditions that were optimal for HIV-RT (6 mM MgCl₂, 100 μM dNTPs, pH=8). Reactions were not very sensitive to salt concentrations showing little difference at 10 vs. 80 mM KCl (data not shown). The addition of “crowding” agents like polyethylene glycol (PEG) did not reproducibly stimulate the reactions. In Figure 2-6, the reactions were performed with 2 nM template and 8 nM primer and 80 nM RT in the presence and absence of NC (4 μM) as indicated. Serial 1:2 dilutions of the reaction performed in the absence of NC were made and these dilutions were loaded in lanes as indicated. In the last lane without NC (undiluted reaction), the amount of reaction material loaded is equal to that loaded in the adjacent lane, which is the reaction with NC. Fully extended products were only observed with NC, although there was about an 80-90 % reduction in total primer extension in these reactions. The total primer extension in the reaction with NC can be compared to 1/8th of the reaction without NC. Figure 2-7 shows an enzyme titration in the presence and absence of NC (4 μM). In the presence of NC, inhibition of primer extension was observed along with a significant increase in the proportion of full-length DNA products

when higher enzyme concentrations were used (80, 160, and 320 nM in the 3 lanes on the right). Full-length DNA products were also produced in the absence of NC at the two highest enzyme concentrations (160 and 320 nM), although the proportion of these long products was very low. Increasing the amount of template in these reactions modestly increased the level of primer extension but the effect was much less dramatic than with the smaller template, and even at 16 nM template and 4 nM primer, only a fraction of the primers were extended in reactions with NC (data not shown). Reactions with the longer template were also less reproducible as not all preparations of the RNA yielded long products. Attempts to synthesize even longer products were made using a 7 kb HIV-derived RNA. Inhibition of primer extension by NC was observed in these reactions but no full-length products were made. Attempts were made to synthesize RNA from the HIV insert cloned into pBR322dR downstream of the T7 promoter for use as template in reverse transcription assays. The pBR322dR is the pBR322 plasmid having a T7 promoter, HDV ribozyme and T7 terminators (Figure 2-8). The construct was first digested with restriction enzyme AseI that has a recognition site downstream of the ribozyme sequence and a run-off transcription was performed using T7 RNA polymerase. Full-length DNA synthesis products would have been about 5 kb with this template. Similar to the 7 kb template, inhibition of extension but no full length DNA products were observed.

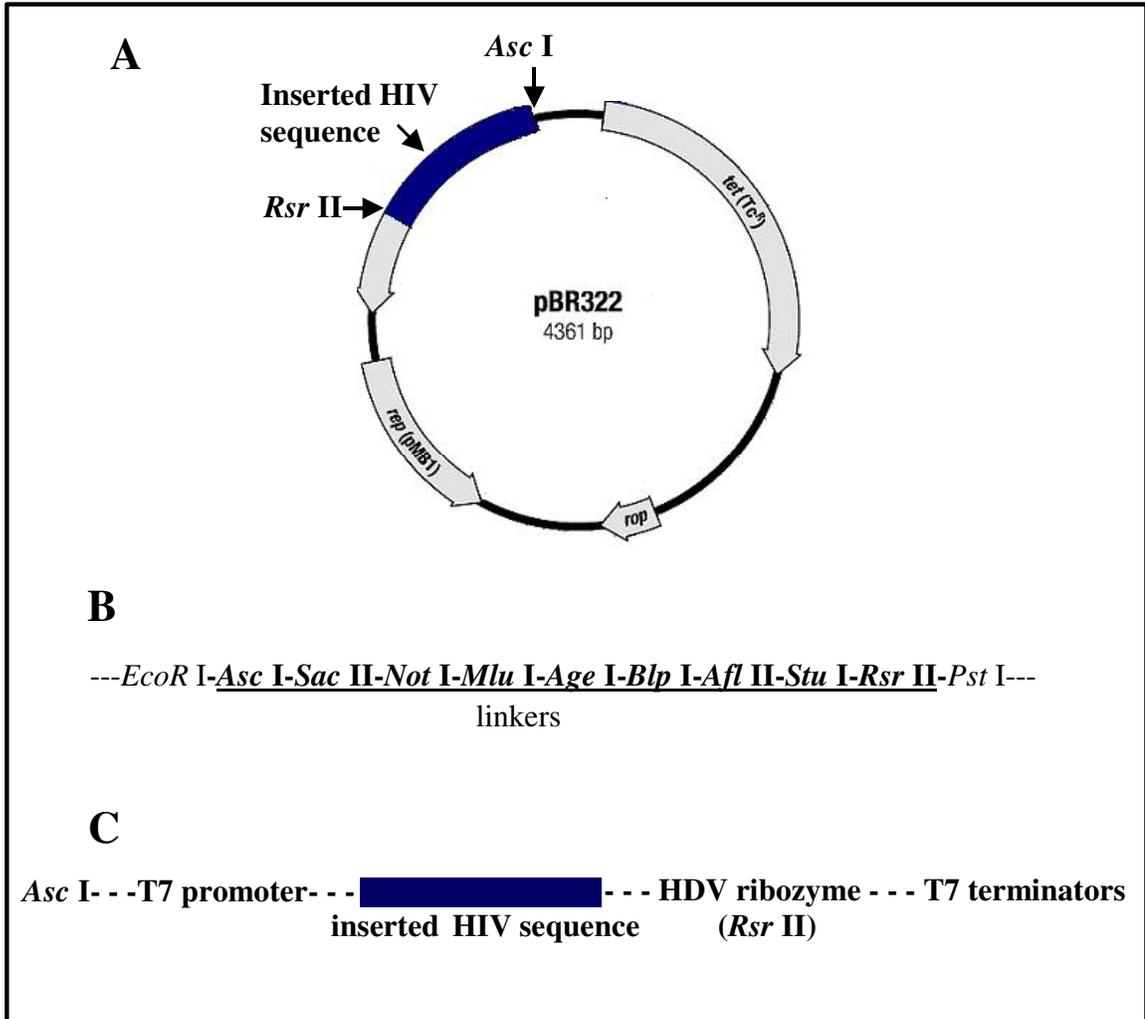


Figure 2-8: pBR322dR is the pBR322 plasmid having a T7 promoter, HDV ribozyme and T7 terminators

Shown in A is pBR322dR which is the pBR322 plasmid having a T7 promoter, HDV ribozyme and T7 terminators and the HIV sequence inserted into the vector at *Asc I* and *Rsr II* sites. Shown in B is the linker region of pBR322. C shows the position of the inserted HIV sequence with respect to the hepatitis delta virus (HDV) ribozyme and two tandem T7 terminators in pBR322dR.

2.5 Discussion

In this report *in vitro* reverse transcription reactions capable of efficiently producing single stranded DNA up to 4 kb from genomic RNA of HIV were characterized. The reactions required large amounts of RT (at least 80 nM) and enough NC to coat the template RNA (Figure 2-2, 2-3, 2-6 and 2-7). NC clearly increased the proportion of full-length DNA products (Figure 2-2, 2-6 & 2-7). This effect of NC is consistent with data from a previous report using a smaller (874 nucleotide) template (209), and was observed only at high NC concentrations, i.e. 4 μ M or higher. At these concentrations there is enough NC to completely coat all the nucleic acids in the reaction at approximately 1 NC per 7 nucleotides. This is typically observed in the HIV virion with the ribonucleoprotein complex consisting of the dimeric RNA genome in association with 2000 to 3000 molecules of NC protein (117). Inhibition of total DNA synthesis was also observed with high NC concentrations, again consistent with previous reports (167, 209). Very high concentrations of NC were however inhibitory to DNA synthesis *in vitro*. This could be because of formation of precipitates rather than functional aggregates at high concentrations of NC. The time course assay (Figure 2-4) shows that it takes at least 45 minutes for the appearance of full-length products *in vitro*. This is because RT is a very sluggish enzyme incorporating 1 to 1.5 nucleotides per second *in vitro* which is approximately 1/10th the rate of eucaryotic DNA polymerases (104, 214-217). A relatively long time is required to generate retroviral DNA *in vivo* (~ 4 hours from infection to the appearance of a 9.7 kb DNA) and about 8-12 hour for the completion of reverse transcription (218). Using PCR analysis of products produced during reverse transcription a rate of about 150-180 nucleotides/min was calculated during infection of

cell with HIV (213). This rate is about 5 times as fast as the rate in the *in vitro* reactions performed here. The reason for the discrepancy is not clear but there are a number of possibilities. First, the cellular rate seems quite high in comparison to the times estimated to complete proviruses by others. The four hour completion time stated above would suggest a rate of about 83 nucleotides/min based on 20,000 base incorporated for the plus and minus strands together. Also, since strand transfer is likely to be slower than continuous synthesis, the high rate of transfer *in vitro* may make the process somewhat slower (see Chapter 5).

Since more primers are extended in reactions without NC this suggests that NC somehow destabilizes primer-template interactions, a finding that is consistent with NC's helix destabilizing activity (129, 149, 163). Consistent with these findings using longer primers (50 rather than 20 nucleotides) improved primer utilization and adding a large excess of template also improved primer utilization (Figure 2-5). Another explanation for decrease in primer extension in the presence of NC is that NC may not completely coat all the nucleic acids in the reaction pool *in vitro* even when added at concentrations enough to completely coat all nucleic acids. Synthesis to the end of the RNA template may occur only if these templates are completely coated with NC as in case of the virion where the genomic RNA exists as a ribonucleoprotein complex. NC bound to nucleic acids may also help preserve the replication intermediate and help prevent RT from falling off the growing DNA thereby promoting synthesis to the end of the template. In Figure 2-6 full-length DNA products are observed at high enzyme concentrations even in the absence of NC. This could be because at very high enzyme concentrations, there are

many more molecules of enzyme in the vicinity of the growing DNA molecule. Hence even if RT falls off the replication complex, RT molecules in close proximity can quickly bind to the RNA template thereby allowing synthesis to proceed to the end of the RNA template explaining why fully extended DNA products are seen at high enzyme concentrations in the absence of NC.

Interestingly, reactions with the 4 kb template showed a much higher level of inhibition than the shorter template (Figure 2-2 and 2-7), while attempts to synthesize even longer products (7 kb) failed. The 4 kb template is essentially a longer version of the 1.9 kb template that extends further toward the 3' end of the HIV genome. It was not clear whether the lower efficiency resulted from the increased length or a different priming position. However, length does not appear to be the only factor in the efficiency of primer extension in the reactions. The much smaller template used in the previous report noted above was also strongly inhibited by NC (209). Taken together the results suggest that a combination of template length and primer location/sequence may be important in determining how efficient primer extension is in the presence of NC. Note also that the efficiency of synthesizing full-length RNA transcripts *in vitro* decreases with increasing size (219). Therefore for the 4 and 7 kb transcripts a large proportion of the RNA likely consists of incompletely synthesized strands lacking the primer binding site. How this would affect the reverse transcription reaction is unclear.

An attempt was made to synthesize RNA from the HIV sequence cloned into pBR322dR downstream of the T7 promoter. This approach was tried as run-off

transcription often gives 3' heterogeneous ends by adding additional non-template-directed bases to the 3' end. Also, RNA polymerases have a tendency to fall-off the template before reaching the end of the DNA especially with longer templates (220-222). A self-cleaving ribozyme sequence near the 3' end of the linearized template self cleaves the 3' end of newly formed transcripts producing RNA with homogeneous 3' ends (223). Although RNA was successfully produced from this approach, no fully extended DNA products (5 kb) were observed with or without NC. It was not clear whether this was due to length restrictions for long DNA synthesis or the integrity of the RNA. Although this approach generates unique 3' ends if RNA polymerase synthesized through the ribozyme region, polymerase that terminate before this point can still produce smaller RNAs that could interfere with synthesis.

Chapter 3 Role of high molecular weight aggregate complexes in the synthesis of long reverse transcription products from genomic RNA segments of HIV

3.1 Introduction

HIV nucleocapsid protein (NC) is known to play an important role in several crucial steps of reverse transcription. These roles of NC can be credited to its nucleic acid chaperone activity. It has been shown that RT pauses at discrete stem structures in the genomic template during minus strand DNA synthesis (184, 224, 225). NC has been shown to melt such structures (also called NC's helix destabilizing activity) of RNA allowing synthesis to continue. Therefore, NC affects RT cDNA synthesis. *In vitro*, NC has been shown to increase the proportion of long cDNA transcripts from genomic RNA produced by reverse transcriptase (209). In addition to the helix destabilizing activity of NC described above, NC has the ability to aggregate nucleic acids (121, 168-171). The NC-induced aggregation of single-stranded RNA was studied using quasielastic dynamic light scattering and optical density measurements (170) and also with electron microscopy (169). The ordered growth of large nucleic acid-NC aggregates was found to be independent of the length and sequence of RNA molecules (169). These aggregates grew with time by fusion of smaller aggregates to give larger aggregates similar to the Ostwald ripening mechanism.

The ability of NC to aggregate nucleic acids promotes the nucleation step of the annealing reaction. Nucleation is a diffusion-limited association and is often retarded by

electrostatic repulsion between annealing strands and also by the low probability of correct positioning of nucleotides for annealing. The nucleic acid aggregation by NC facilitates attraction between nucleic acids strands (121, 168-171) and complementary sequences can search for each other within the aggregate. NC is highly cationic, with 15 positively charged amino acids distributed throughout its 55 amino acids. Hence NC does not self-aggregate even at high concentrations and is known to bind nucleic acids non-cooperatively (139, 140, 208, 226). Various reports suggest that NC is highly mobile when bound to nucleic acids (154, 227-231). This mobility may be an important feature for efficient aggregation as observed with multivalent cationic ligands, which are aggregating agents in their nucleic acid-bound state (232-235). Therefore, NC-induced nucleic acid aggregation appears to be facilitated by polyelectrolyte attraction similar to that observed for multivalent cations (235). The main aggregating ability of NC was mapped to its N-terminal 3_{10} helix (68, 119, 121, 169, 170). The role of zinc fingers in the aggregation of nucleic acids was studied using fingerless NC mutants (171). It was found that efficient aggregation of nucleic acids occurred only at very high concentrations of the NC mutants suggesting that these fingers were important for effective aggregation.

Various reports have demonstrated that under the right conditions, NC can stimulate the rapid formation of large aggregated complexes that contain NC, RT and RNA (167, 236). These complexes are functional as isolated complexes can synthesize DNA in the presence of dNTPs, use tRNAs to prime viral RNA synthesis and catalyze minus strand strong-stop strand transfers (236). Such complexes also protect viral RNA against nuclease degradation (236). NC within these complexes directly interacts with RT

and plays an important role in recruiting RT into the complexes during viral DNA synthesis (167). Experiments conducted in this section, show that such high molecular weight aggregate complexes are formed during synthesis of long DNA products and that these aggregates play an important role in the synthesis of long DNAs *in vitro*. Also, reverse transcription assays using the long RNAs was conducted with NC finger mutants, 1.1, 2.2 and 2.1 to show that the aggregation activity of NC is important for the synthesis of long DNAs *in vitro*.

3.2 Materials

Plasmid pBKBH10S was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH, from Dr. John Rossi. This plasmid contains an 8.9 kb SstI fragment (nt 222-9154 of the RNA genome) from HIV-1 BH10 inserted into the SstI site. The fragment has all HIV-1 gene coding regions but does not contain the HIV-1 LTR (210). Primers used to prime templates in reverse transcription assays were obtained from Integrated DNA Technologies, Inc. The HIV-RT clone was a generous gift from Dr. Samuel H. Wilson (National Institute of Environmental Health Sciences, Research Triangle Park, NC). HIV-RT was purified according to the protocol described (211). The protein was purified to homogeneity and the purity of the protein was evaluated using Coomassie Blue staining of 10 % SDS-PAGE gels (212). The subunits p51 and p66 of RT were in a 1:1 ratio. Aliquots of HIV RT were stored frozen at -80°C and fresh aliquots were used for each experiment. The HIV NC clone was a generous gift from Dr. Charles McHenry (University of Colorado). NC was purified to apparent homogeneity (as judged from Coomassie Blue staining of 17.5% SDS-PAGE gels (212)) according to the protocol described (123). Quantification was by absorbance at 280 nm using a molar extinction coefficient of 8350 cm⁻¹ M⁻¹ (123). Aliquots of NC were stored frozen at -80 °C, and fresh aliquots were used for each experiment. NC finger mutants 1.1, 2.2 and 2.1 were a gift from Dr. Robert Gorelick (SAIC, Frederick, MD). These proteins were expressed and purified as described (237), and quantified by amino acid analysis on a Beckman Systems 6300 amino acid analyzer (Beckman Coulter, Inc., Fullerton, CA). T7 RNA polymerase, DNase I-RNase-free, and RNase-DNase-free were from Roche Diagnostics. RNase inhibitor was from Promega. T4 polynucleotide kinase

and restriction enzyme HincII was obtained from New England Biolabs. Proteinase K was obtained from Eastman Kodak Co. Radiolabeled compounds were obtained from Amersham. Sephadex G-25 spin columns were from Amika Corp. RNA cleanup kit was from Qiagen. All other chemicals were from Sigma or Fisher Scientific.

3.3 Methods

Preparation of RNA substrates- RNA of approximately 1.9 kb was made by first digesting pBKBH10S with restriction enzyme HincII. The digest was then extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with ethanol. Run-off transcription (performed according to the enzyme manufacturer's protocol) was then conducted using 5 µg of the digest plasmid and T7 RNA polymerase enzyme to generate 1.9 kb RNA. The transcription reactions were treated with 2 µl of 10 units/µl of DNase I-RNase-free enzyme for 15 min to digest away the template DNA. The RNA was purified using the Qiagen RNA cleanup kit. The amount of recovered RNA was determined spectrophotometrically from optical density. The integrity of the RNA was checked on a 1% agarose gel as described before.

RNA-DNA Hybridization- DNA primer 5'-CTGAAGCTCTCTTCTGGTGG-3' that bound specifically to the 1.9 kb RNA template was ³²P-labeled at the 5'-end with T4 polynucleotide kinase according to the manufacturer's protocol. The RNA template was hybridized to the complementary labeled primer by mixing primer: transcript at a ~ 1:1 ratio in 50 mM Tris-HCl (pH 8.0), 1mM dithiothreitol and 80 mM KCl. The mixture was heated to 70°C for 5 min and then slowly cooled to room temperature.

Reverse transcription reactions with or without NC mutants- RNA template-DNA primer hybrids (4 nM final concentration of RNA) were pre-incubated for 5 min along with additional template RNA (12 nM) in the presence or absence of the NC mutants (4 µM) in 21 µl of buffer (see below) at 37°C. The reactions were initiated by addition of 4

μ l of HIV-RT (80 nM final in reactions). The following reagents at the indicated final concentrations were also included in the reaction mixtures: 50 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 80 mM KCl, 6 mM MgCl₂, 100 μ M dNTPs, 5 mM AMP (pH 7.0), 25 μ M ZnCl₂ and 0.2 units/ μ l RNase inhibitor. Reactions were allowed to incubate for 75 min. The reactions were stopped by adding 2 μ l of a solution containing 250 mM EDTA (pH 8.0) and 5 ng of RNase-DNase-free enzyme and allowed to digest for 20 min at 37°C. Nine μ l of proteinase K at 2 mg/ml in 1.25 % SDS, 15 mM EDTA (pH 8.0) and 10 mM Tris (pH 8.0) was then added to the above mixture, which was placed at 65°C for 1 hour. Finally 7 μ l of 6X alkaline dye (300 mM NaOH, 6 mM EDTA, 15% glycerol, 0.15% bromophenol blue) was added to the mixture and the samples were resolved on 1% alkaline agarose gel containing 50 mM NaOH and 1 mM EDTA (pH=8). Extended DNA products were observed using a Bio-Rad Molecular Imager FX.

Experiments testing aggregate formation- Reactions were set up and initiated similar to the reverse transcription experiment described above except that the final reaction volume was 50 μ l. As indicated in Figure 3-1 and 3-2, the reactions were centrifuged for 1 min at 12,000 x g at the end of 60 min and at the end of 2, 10 or 60 min respectively after initiating with RT. The pellet and the supernatant fractions from the 2 and 10 min samples were then subjected to reverse transcription (58 and 50 min, respectively) after adding back the reverse transcription buffer described above to the pellet fractions only. Pellet and supernatant fractions from all reactions were stopped and treated with proteinase K and samples were resolved on alkaline agarose gels as described above.

Reverse transcription assay with excess template- Reactions were set up and initiated similar to the reverse transcription experiment described above. In one set of reactions, RNA template-DNA primer hybrids (4 nM final concentration of RNA) were pre-incubated for 5 min along with additional template RNA (12 nM) in the presence or absence of the NC (4 μ M) in 21 μ l of buffer (described above) at 37°C and in the other set of reactions, additional template RNA (12 nM and 28 nM, as indicated) was added after adding NC to the primer-template hybrid. Reactions were carried out as described above and extended DNA products were resolved on 1% alkaline agarose gels observed using a Bio-Rad Molecular Imager FX.

Gel electrophoresis- One percent alkaline agarose gels containing 50 mM NaOH and 1 mM EDTA (pH=8), 1 % native agarose gels in Tris-Borate-EDTA buffer were prepared and subjected to electrophoresis as described (212).

3.4 Results

Long DNA products in vitro are synthesized in a high molecular weight aggregate that forms rapidly in the presence of NC and contains NC, RT, primer, and RNA- Stable HIV-1 nucleocapsid complexes can be generated *in vitro* when NC protein is added to RNA (169, 170, 236). The NC within these nucleoprotein complexes recruits reverse transcriptase into the complexes during viral DNA synthesis through NC-RT interactions (167). The complexes were large and could be pelleted by slow speed centrifugation and were found to be competent for DNA synthesis (236). To test for the formation of aggregates in the reactions that produce long DNAs, the assay was performed using the 1.9 kb RNA as template in the presence or absence of NC. After 1 hour the material was centrifuged for 1 min at 12,000 x g in a microfuge. The pellet and supernatant fractions were then run on a 1 % alkaline agarose gel. The pellet fraction of the reactions with NC had the full-length DNA while the supernatant fraction showed no products (Figure 3-1). Some smaller products were apparent in the supernatant fractions using darker exposures. All products in reactions without NC were short and contained in the supernatant. The results suggest that high molecular weight aggregates are formed during synthesis of long DNA products.

NC is known to form precipitates at high concentrations so the high molecular weight pellet fractions could be a result of NC precipitation rather than a discrete high molecular weight functional aggregate. In order to determine whether these high molecular weight aggregates were functional and promoted synthesis of long DNA, the

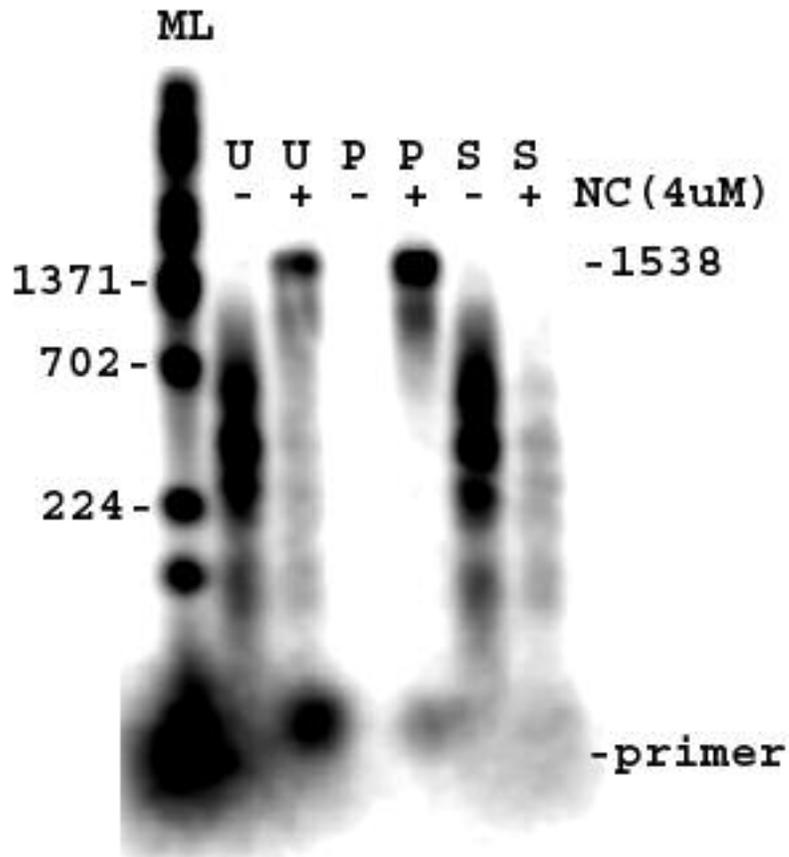


Figure 3-1: Reverse transcription assay showing formation of high molecular weight aggregates.

The 1.9 kb RNA was used as template. Lanes with reactions with NC (4 μ M) are denoted as + and those without NC as -. U denotes lanes with uncentrifuged control reactions. Pellet (P) and supernatant (S) fractions of the reactions with and without NC were obtained after centrifuging reactions at the end of one hour for 1 min at 12,000 x g as described in "Methods". ML denotes lane with molecular marker (in nucleotides).

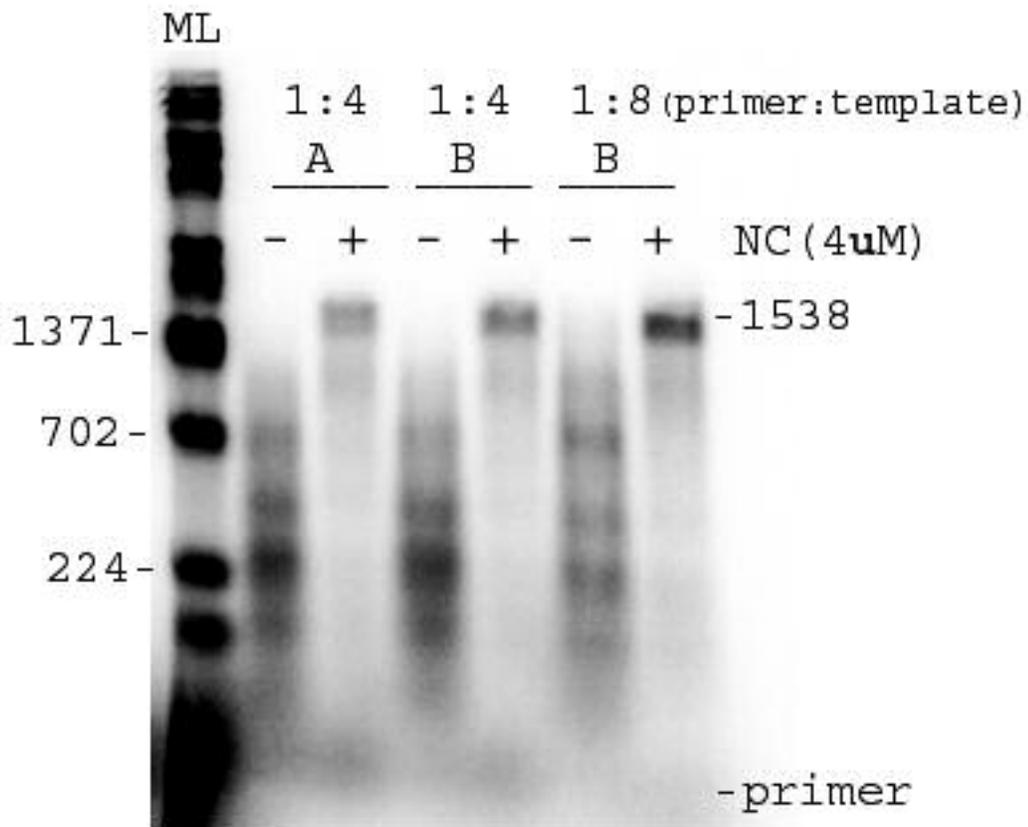


Figure 3-3: Reverse transcription reaction to show that inefficient primer-templates binding in reactions with NC

The 1.9 kb RNA was used as template. Lanes with reactions with NC (4 μ M) are denoted as + and those without NC as -. In all reactions primers were hybridized to templates in a 1:1 ratio (4 nM primer: 4 nM template). In reactions marked A, 3-fold excess template (12 nM) was added before adding NC to the reaction and in reactions marked B, 3-fold (12 nM) and 7-fold (21 nM) excess template were added after adding NC such that the ratio of primer to template in these reactions is 1:4 and 1:8 respectively. ML denotes lane with molecular marker (in nucleotides).

reactions were centrifuged early (2 and 10 min of Figure 3-2) after the start of reverse transcription and pellet and supernatant fractions were incubated for 58 and 50 min respectively, at 37°C. In these experiments the pellet was resuspended in buffer containing divalent cation and dNTPs. In each case, the pellet fractions of the reactions with NC had the larger products and the supernatant fractions showed no DNA products. Once again darker exposures did reveal some small products in the supernatants. The results show that aggregates are formed early in the reactions. The full-length DNA products are not formed early in the reverse transcription. Results from time-course experiments have shown that it takes at least 45 min for full-length DNA to be generated, suggesting that these aggregates are functional and play an important role in the synthesis of long DNA products *in vitro*. Figure 3-3 is an experiment done to compare DNA synthesis when excess template is added before and after addition of NC. The addition of excess template improves total DNA synthesis irrespective of whether the excess template is added before or after pre-incubation of the reaction with NC. This suggests the primers in the reaction bound inefficiently to templates in the aggregates and excess template is required to get more primers bound and extended.

NC finger mutants lacking either finger 1 or 2 or switching their positions also stimulate the synthesis of long DNA products in vitro- HIV-NC has two non-identical zinc fingers, an N- and a C- terminal finger, denoted 1 and 2, respectively. Three NC mutants, 1.1 NC, 2.2 NC and 2.1 NC were used. In mutant 1.1, finger 1 replaces finger 2 giving the protein two copies of finger 1. In 2.2, finger 2 replaces finger 1 giving this protein two copies of finger 2 and 2.1 NC is a finger switch mutant in which the positions

of the zinc fingers are switched. It was previously reported that the N- and C-terminal zinc fingers of NC are not biologically equivalent (134, 135, 238). Previous work from several laboratories has shown that the two fingers possess different functional activities with finger 1 being more important for helix-destabilizing activity than finger 2 (135, 238). Results showed that 1.1 and 2.1 retained helix-destabilizing activity while 2.2 had little. In contrast, all the mutants were able to stimulate the annealing of non-structured complements, suggesting that they retained “aggregating/condensing” activity that is required to bring nucleic acids into close proximity (135). To determine if the finger mutants could stimulate the production of long DNAs, an NC titration was performed as described for wild-type NC using the 1.9 kb RNA as template, and the DNA products were run on a 1% alkaline agarose gel (Figure 3-4). All three NC finger mutants increased the proportion of full-length DNA products similar to wild-type NC. These results are consistent with data from the previous report using the 874 nucleotide template (209). Since results with 2.2, which has little unwinding activity, were similar to those with wild type, this suggests that the unwinding activity of NC may not play a role in the synthesis of long DNAs *in vitro*.

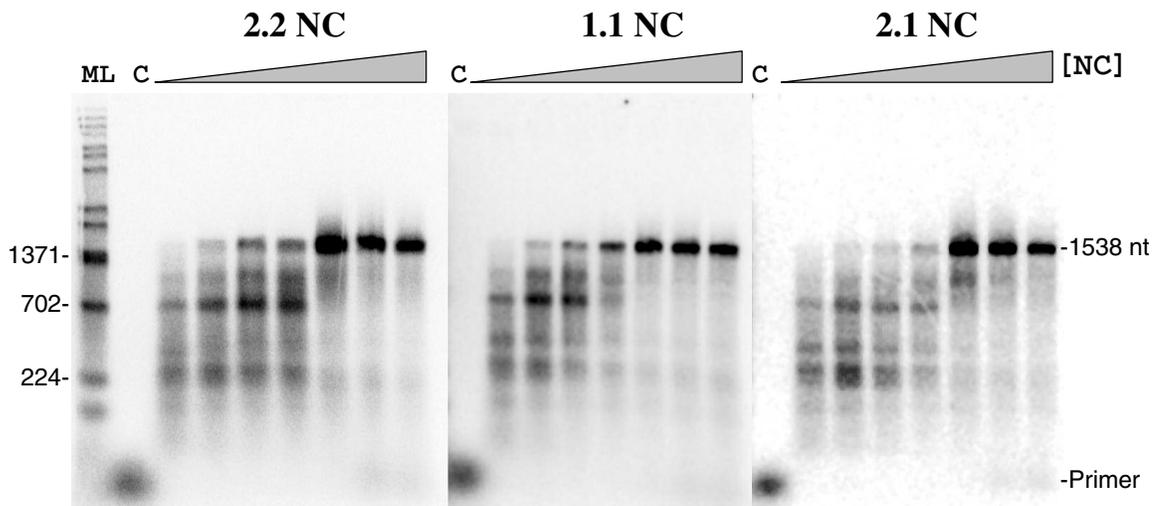


Figure 3-4: Reverse transcription assay with increasing concentrations of NC finger mutants.

Shown are autoradiograms of assays with 1.9 kb RNA template with increasing concentrations (left to right, 0, 0.5, 1, 2, 4, 6, 8 μ M) of NC finger mutants 2.2, 1.1 and 2.1 (as indicated). See “Results” for description of NC mutants. ML denotes lane with molecular marker (in nucleotides).

3.4 Discussion

In this report *in vitro* synthesis of long reverse transcription products from genomic RNA of HIV in high molecular weight aggregates is described. The aggregates formed rapidly and were pelleted by low speed centrifugation indicating that they were large (Figure 3-1, 3-2). The aggregates were functional and synthesize long DNAs in the presence of RT and NC when supplemented with dNTPs and Mg^{2+} . They were consistent with previously reported NC aggregates that were able to carry out reverse transcription (see Introduction), but were not examined for the ability to produce long DNAs. A possible explanation for the large products made in aggregates is that RNA molecules would be in close proximity to each other thereby allowing strand transfers to occur more easily. The nucleic acid aggregation facilitated by NC may promote attraction between nucleic acids strands such that complementary sequences can find each other within the aggregate (121, 168-171). This could further be facilitated by the high local concentrations of NC and RT in the complexes. Aggregates may thereby promote the synthesis of long DNA products by concentrating the nucleic acids, RT and NC into a smaller area perhaps mimicking the role of the capsid environment within the cytoplasm of the host cell.

An increase in synthesis of long DNA products was observed by addition of excess template either before or after aggregate formation. This could be because primers in the reaction bound inefficiently to templates in the aggregates. Hence more products are formed when excess template is added as more primers are bound and extended. If the primers were non-extendable or could not access RT in the aggregate then adding more

templates would not lead to extension. Since more primers are extended in reactions without NC this suggests that NC somehow destabilizes primer-template interactions, a finding that is consistent with NC's helix destabilizing activity (129, 149, 163).

Synthesis of long DNA products *in vitro* did not seem to require NC's helix destabilizing activity as a mutant NC with very low activity (Figure 3-4, 2.2 NC) was as effective as wild type in the assays. This suggests that the aggregation/condensation activity of NC may be all that is required for synthesis of long DNA products *in vitro*. This activity is likely responsible for the large aggregates observed in reactions with NC. Although we do not have definitive evidence showing that long products can only be synthesized in aggregates, that conclusion would be consistent with the results. Aggregated complexes isolated just two minutes into the reactions were capable of long DNA synthesis and long products were only associated with material in the reactions that pelleted with slow speed centrifugation (Figure 3-2).

Chapter 4 Role of dimerization region of HIV and processivity of RT in the synthesis of long reverse transcription products from genomic RNA segments of HIV *in vitro*

4.1 Introduction

HIV, like all retroviruses contain two copies of genomic RNA, which are non-covalently linked in an apparent parallel orientation close to their 5' ends by a structure called the Dimer Linkage Structure (DLS) or the Dimer Initiation Sequence (DIS). The dimerization initiation site (DIS) of HIV RNA is a hairpin structure that contains in the loop a 6-nucleotide self-complementary sequence flanked by two 5' and one 3' purines (see Chapter 1, Figure 1-2). The self-complementary sequence, as well as the flanking purines are critical for dimerization of HIV RNA, which is mediated by formation of a "kissing-loop" complex between the DIS of each monomer (239). The nucleotides of the sequence that constitute the kissing loop of one monomer recognize the complementary nucleotides constituting the kissing loop of the other monomer. This loop-loop interaction initiates the dimerization process and is therefore called the DIS (240). Upon recognition of another DIS, the base pairs of the monomer split and the equilibrium shifts towards the formation of dimers thereby reducing the activation energy of the monomer-dimer conformational switch (241). The interaction and base pairing between the two loops may induce subsequent annealing between the two stems and other downstream sequences (see Figure 1-2) (241). Dimerization is known to play an important role in preferential packaging (or encapsidation) of two genomic RNAs within the capsid (242) and is essential for stabilization of the genome. The dimerization of the two RNAs holds them

in close proximity (243), which could potentially allow rapid strand transfers between the two RNAs. Hence dimerization may play an important role in recombination (86, 244). It has also been shown that DIS is important in mediating the complete synthesis of viral cDNA in infected cells (245). In this section, RNA templates lacking the DIS were used in reverse transcription assays. These assays showed that this region was not important in the synthesis of long cDNA products from genomic RNA of HIV in an *in vitro* system.

HIV Reverse Transcriptase (RT) is a moderately processive enzyme (see Chapter 1). Processivity of a polymerase is defined as the average number of nucleotides the enzyme adds to the growing chain in a single binding event with the primer-template. For HIV-RT, a processivity of approximately 100 nucleotides has been estimated *in vitro* (214, 218). Therefore, several rebinding events would be required to complete synthesis of the approximately 10 kb provirus DNA. *In vitro*, RT generally falls off the template at discrete locations that are referred to as pause sites. These are usually particular sequences that the enzyme has difficulty traversing (184, 246) or more commonly, secondary structures on the genome (247) that impede the enzyme's progress. HIV NC has been shown to have a modest effect on processivity (166, 204) by helping to melt out secondary structures. NC is also known to enhance processivity of reverse transcription by promoting RT-catalyzed strand transfer reactions through modulation of RNase H activity (159, 248, 249). Studies by Tanchou et al (236) have shown that reverse transcription within nucleoprotein complexes generated *in vitro* appeared to be more processive than with viral RNA alone. Therefore it is possible that enhanced RT processivity within the aggregates described in Chapter 3 could play a role in the

generation of long products. In this section, reverse transcription assays showed that enhanced processivity does not play a role in synthesis of long DNA products *in vitro*.

4.2 Materials

Plasmid pBKBH10S was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH, from Dr. John Rossi. This plasmid contains an 8.9 kb SstI fragment (nt 222-9154 of the RNA genome) from HIV-1 BH10 inserted into the SstI site. The fragment has all HIV-1 gene coding regions but does not contain the HIV-1 LTR (210). PCR primers and primers used to prime templates in reverse transcription assays were obtained from Integrated DNA Technologies, Inc. The HIV-RT clone was a generous gift from Dr. Samuel H. Wilson (National Institute of Environmental Health Sciences, Research Triangle Park, NC). HIV-RT was purified according to the protocol described (211). The protein was purified to homogeneity and the purity of the protein was evaluated using Coomassie Blue staining of 10 % SDS-PAGE gels (212). The subunits p51 and p66 of RT were in a 1:1 ratio. Aliquots of HIV RT were stored frozen at -80°C and fresh aliquots were used for each experiment. The HIV NC clone was a generous gift from Dr. Charles McHenry (University of Colorado). NC was purified to apparent homogeneity (as judged from Coomassie Blue staining of 17.5% SDS-PAGE gels (212)) according to the protocol described (123). Quantification was by absorbance at 280 nm using a molar extinction coefficient of 8350 cm⁻¹ M⁻¹ (123). Aliquots of NC were stored frozen at -80 °C, and fresh aliquots were used for each experiment. *Taq* polymerase was from Eppendorf. T7 RNA polymerase, SP6 RNA polymerase, DNase I-RNase-free and RNase-DNase-free were from Roche Diagnostics. RNase inhibitor was from Promega. T4 polynucleotide kinase and Restriction enzyme HincII was obtained from New England Biolabs. Proteinase K was obtained from Eastman Kodak Co. Radiolabeled compounds were obtained from Amersham. Sephadex

G-25 spin columns were from Amika Corp. RNA cleanup kit was from Qiagen. All other chemicals were from Sigma or Fisher Scientific.

4.3 Methods

PCR Amplification of DNA substrates for RNAs without dimerization signal-Two PCR primers, 5'-**GATTAGGTGACACTATAGGAATTAGATCGATGGGAAAA**-3' and 5'-CTGAAGCTCTCTTCTGGTGG-3' were designed to yield RNA templates without dimer initiation site (bases 18 to 135 of the HIV genome insert in pBKBH10S (239-356 of genomic RNA)) and amplified DNA from position 145 to 1538 (genome bases 366-1759) on the HIV insert. An SP6 promoter sequence (in bold) was included on one primer in one of the primers to allow transcription of the DNA by SP6 RNA polymerase. PCR reactions were performed with *Taq* polymerase according to the enzyme manufacturer's protocol using the provided buffer. One hundred pmol of each primer was used. Reactions included 30 cycles of denaturation, annealing and extension at temperatures of 94°C for 1 min, 50°C for 1 min and 72°C for 2 min, respectively followed by one cycle of extension at 72°C for 5 min. The PCR products were run on a 1% agarose gel, extracted by dialysis and purified as described (212), and used to prepare RNA as described below.

Preparation of RNA substrates- RNAs of approximately 1.9 were made by first digesting pBKBH10S with restriction enzyme HincII. The digests were then extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with ethanol. Run-off transcription (performed according to the enzyme manufacturer's protocol) was then conducted using 5 µg of the digest plasmid and T7 RNA polymerase enzyme to generate 1.9 RNAs. Run-off transcription was also performed using ~5 µg of purified PCR DNAs described above and SP6 RNA polymerase to generate RNAs without the dimerization

signal (approximately 1.4 kb). The transcription reactions were treated with 2 μ l of 10 units/ μ l of DNase I-RNase-free enzyme for 15 min to digest away the template DNA. The RNA was purified using the Qiagen RNA cleanup kit. The amount of recovered RNA was determined spectrophotometrically from optical density. The integrity of the RNA was checked on a 1% agarose gel as described before.

RNA-DNA Hybridization- DNA primers that bound specifically to the RNA templates: 5'-CTGAAGCTCTCTTCTGGTGG-3' to the 1.9 kb and 1.4 kb (dimer minus) templates were 32 P-labeled at the 5'-end with T4 polynucleotide kinase according to the manufacturer's protocol. The RNA templates were hybridized to the complementary labeled primer by mixing primer: transcript at a ~ 1:1 ratio in 50 mM Tris-HCl (pH 8.0), 1mM dithiothreitol and 80 mM KCl. The mixture was heated to 70°C for 5 min and then slowly cooled to room temperature.

Reverse transcription reactions using dimer minus RNA with or without NC- RNA template (dimer minus)-DNA primer hybrids (4 nM final concentration of RNA) were pre-incubated for 5 min along with additional template RNA (12 nM) in the presence of increasing concentrations of NC (0.5, 1, 2, 4, 6 and 8 μ M) or absence of NC (0 μ M) in 21 μ l of buffer (see below) at 37°C. The reactions were initiated by addition of 4 μ l of HIV-RT (80 nM final in reactions). The following reagents at the indicated final concentrations were also included in the reaction mixtures: 50 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 80 mM KCl, 6 mM MgCl₂, 100 μ M dNTPs, 5 mM AMP (pH 7.0), 25 μ M ZnCl₂ and 0.2 units/ μ l RNase inhibitor. Reactions were allowed to incubate for 75

min. In some reactions the amounts of excess template, RT, or NC were varied as indicated. The reactions were stopped by adding 2 μ l of a solution containing 250 mM EDTA (pH 8.0) and 5 ng of RNase-DNase-free enzyme and allowed to digest for 20 min at 37°C. Nine μ l of proteinase K at 2 mg/ml in 1.25 % SDS, 15 mM EDTA (pH 8.0) and 10 mM Tris (pH 8.0) was then added to the above mixture, which was placed at 65°C for 1 hour. Finally 7 μ l of 6X alkaline dye (300 mM NaOH, 6 mM EDTA, 15% glycerol, 0.15% bromophenol blue) was added to the mixture and the samples were resolved on 1% alkaline agarose gel containing 50 mM NaOH and 1 mM EDTA (pH=8). Extended DNA products were observed using a Bio-Rad Molecular Imager FX.

Experiments testing processivity of HIV-RT- RNA template (1.9 kb)-DNA primer hybrids (4 nM final concentration of RNA) were preincubated for 5 min at 37°C with 3-fold excess (12 nM final concentration) RNA template and 4 μ M final concentration of NC in 17 μ l of buffer containing 50 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 80 mM KCl, 5 mM AMP (pH 7.0), 25 μ M ZnCl₂ and 0.2 units/ μ l RNase inhibitor. Four μ l of HIV-RT (80 nM final concentration in reactions) was then added and further incubated for 3 min at 37°C. Reactions were initiated by adding 4 μ l of a supplement containing MgCl₂ and dNTPs in the above buffer such that the final concentrations were 6 mM and 100 μ M, respectively. In the reactions with “trap”, 5 μ g of poly(rA)-oligo(dT)₂₀ (8:1, w/w) was included in the supplement to sequester RT molecules that dissociated from the substrate. In control reactions to test the effectiveness of the trap (see Results), the trap mix was added before adding the enzyme, incubated for 3 min at 37°C and then the reactions were initiated by adding enzyme. Reactions were allowed to incubate at 37°C

for 1 hour. The reactions were stopped and treated with proteinase K as described above. The samples were extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with ethanol. Due to the inhibition by NC of primer extension (described above), ten reactions with NC in the presence of trap were combined. The samples were resuspended in 5 μ l of water and 5 μ l of 2X formamide dye (90% formamide, 10 mM EDTA (pH 8.0), 0.1% xylene cyanol, 0.1% bromphenol blue) was added and the samples were resolved on a 5% denaturing polyacrylamide gel containing 7 M urea. Several 1:2 dilutions of the reactions without NC in presence of trap are shown to make it easier to compare them to the reactions with NC and trap.

Gel electrophoresis- One percent alkaline agarose gels containing 50 mM NaOH and 1 mM EDTA (pH=8), 1 % native agarose gels in Tris-Borate-EDTA buffer and denaturing 5% polyacrylamide gels (19:1) (acrylamide:bisacrylamide), containing 7 M urea were prepared and subjected to electrophoresis as described (212).

4.4 Results

The region of the HIV genome required for dimer formation is not needed for the synthesis of long DNA products in vitro- This experiment was done to determine if the dimer initiation signal (dimer initiation signal/dimer linkage structure) is essential for the formation of long DNA products *in vitro*. The dimer initiation signal is required for dimer formation between the two RNA molecules that make up the HIV genome and hold the RNAs in close proximity (243). This could potentially allow rapid strand transfers between the RNAs which could lead to long products. Both the 1.9 and 4 kb RNAs used in Chapter 2 contained the dimerization region. Therefore, RNA templates that lacked the dimerization region were made (described in Methods) and the reverse transcription assay was carried out. A reaction with the 1.9 kb RNA with a deleted dimerization region (now 1.4 kb) in the presence of increasing amount of NC is shown in Figure 4-1. Long DNA products were synthesized from these RNA templates as efficiently as from those that had the dimerization region (compare Figure 2-2 and 4-1) indicating that the dimerization region does not have a role in the synthesis of long DNA products *in vitro*.

Processivity of RT has no role in the synthesis of long DNA products in vitro- NC has been shown to have a modest effect on the processivity of RT (166, 204). This is generally attributed to melting by NC of some secondary structures in the template. An increase in processivity could also have contributed in producing long products. The following experiment was performed to test this. The 1.9 kb RNA template was used and the reverse transcription assay was carried out in the absence or presence of poly(rA)-oligo(dT) trap. The trap sequesters enzyme molecules that dissociate from the substrate

thereby limiting synthesis to a single binding event between the enzyme and the substrate (250). To test the effectiveness of the trap, it was added to the reaction before the enzyme. After enzyme addition incubation was continued for 1 hour. No significant DNA synthesis products were evident in this reaction (Fig. 4-2, lanes marked as 'C') indicating that the trap sequestered RT over the entire reaction. Assays performed in the absence of NC with trap showed products up to about 800 nucleotides in length. Several dilutions of these reactions (lanes shown as -NC, + trap) are shown to make it easier to compare them to the reactions with NC (lane +NC, + trap). Due to the inhibition by NC of primer extension (described in Chapter 2), several reactions were combined for the sample shown in lane +NC, +trap (see Methods). No notable increase in the average length of products was evident in the reactions with NC. This indicates that there was no significant increase in processivity in the presence of NC and this was not a factor in production of the long DNA synthesis products.

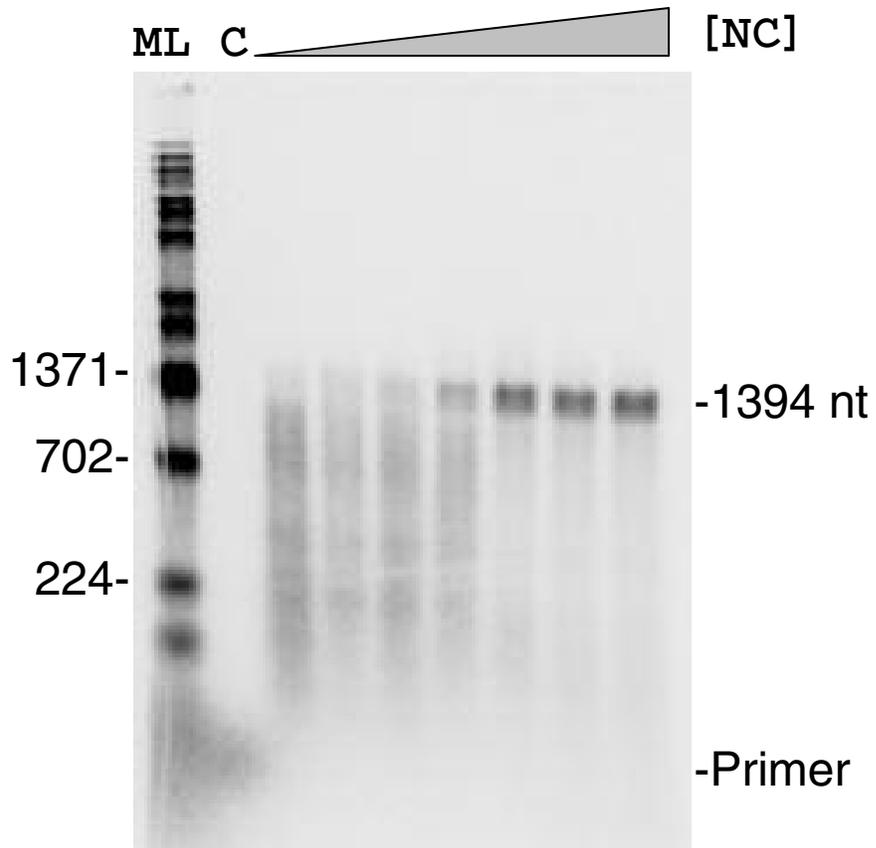


Figure 4-1: Reverse transcription assay with increasing NC concentrations using RNA without dimerization signal as template.

Shown is an autoradiogram of an assay with 1.4 kb RNA (genomic RNA segment without dimerization signal) using increasing NC concentrations (left to right, 0, 0.5, 1, 2, 4, 6 and 8 μ M) performed under conditions as indicated under “Methods”. The position of full length DNA products is indicated (“1394”). C denotes the lane with reaction carried out without RT. ML denotes lane with molecular marker (in nucleotides).

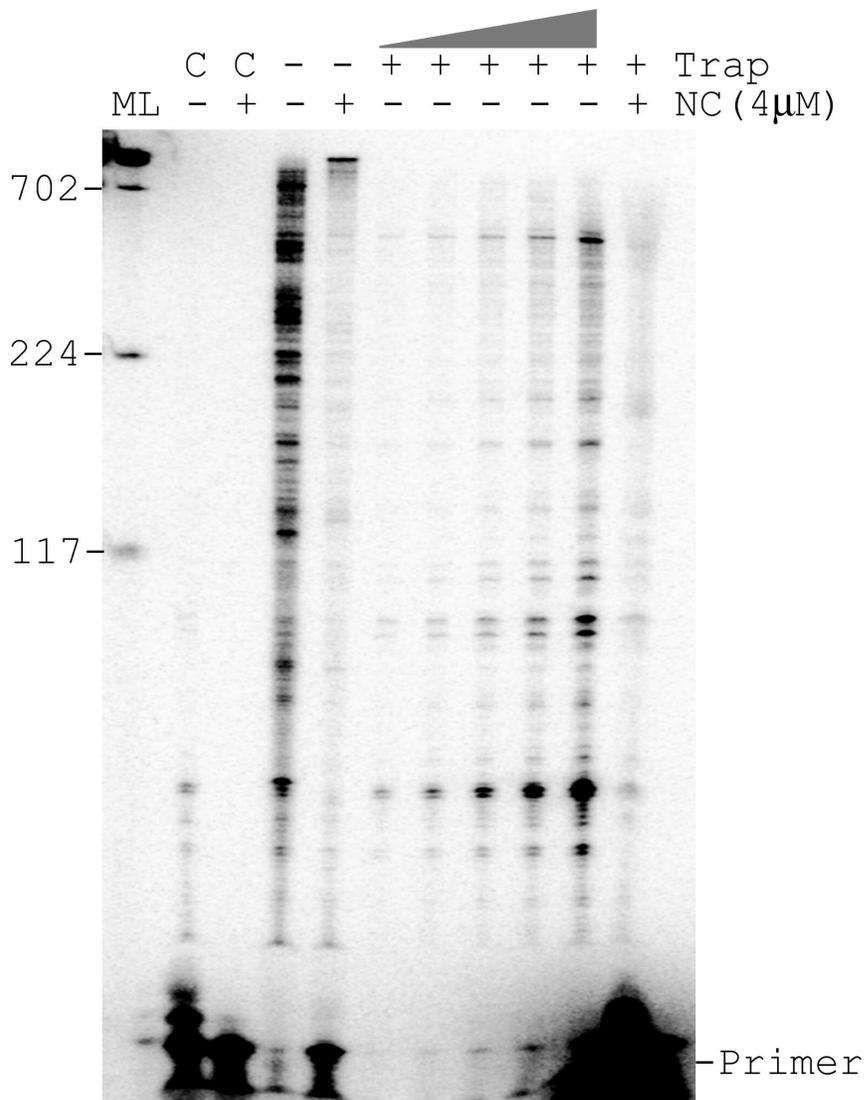


Figure 4-2: Reverse transcription assay to determine role of processivity of RT in the synthesis of long DNA products *in vitro*.

Shown is an autoradiogram of a trap assay using the 1.9 kb RNA segment of HIV as template. Control reactions (“C”) were performed in the presence (+) or absence (-) of 4 μM NC to test the effectiveness of the trap (see Methods). Reactions without trap were also performed with or without NC as indicated. In reactions with poly(rA)-oligo(dT) trap, RT was preincubated with the primer-template in the presence or absence of NC, then initiated by the addition of divalent cation and dNTPs along with trap. The trap sequesters RT molecules that dissociate from the primer-template limiting extension to a single binding event. The trap reactions without NC were serially diluted 1:2 from right to left with the far right lane corresponding to a single reaction and the far left 1/16th of a reaction. The trap reaction with NC is 10 reactions combined after extraction and precipitation, then loaded in a single lane. The samples are run on a 5% polyacrylamide gel. ML denotes lane with molecular marker (in nucleotides).

4.5 Discussion

The experiment testing the role of dimerization of the RNAs in synthesis of long DNAs shows that dimer formation, which holds the two RNA genomes close together, is not required for the formation of long DNA products. Removal of the dimerization region did not affect synthesis of long products (Figure 4-1). This could be because the aggregates formed in the presence of NC (discussed in Chapter 3) may hold the nucleic acids in close proximity to each other thereby creating a concentrated environment required for reverse transcription even without dimerization of the RNAs. Note that no experiments were performed to determine if dimers formed even with the RNAs that contained the dimerization signal. Dimer formation is generally analyzed using nucleic acid segments smaller than those used in these experiments and dimerization is difficult to test with long RNAs.. Dimer formation has been shown to enhance strand transfer in *in vitro* reactions with relatively small nucleic acid substrates (251, 252). The results here show that dimerization is not required to get long products.

From processivity experiments (Figure 4-2), it was clear that enhanced processivity of RT is not required or does not have any role in the synthesis of long DNA products. Hence, for synthesis to proceed to the end of these RNA templates, the RT must rebind several times after dissociation. The high local concentration of contents in the aggregate (discussed in Chapter 3) may also explain why higher processivity of RT is not required or does not have any role in the synthesis of long DNA products. It is important to note that the assays were conducted with poly(rA)-oligo(dT) trap and it is not clear how this may have affected synthesis of long DNAs in aggregates. It is possible that some factor

required for long DNA production was altered by the trap and therefore enhanced processivity cannot be completely ruled out.

Chapter 5 Role of strand transfer in the synthesis of long reverse transcription products from genomic RNA segments of HIV *in vitro*

5.1 Introduction

Strand transfer, also referred to as template switching or strand jumping is a process by which the nascent DNA that is elongated on one RNA template transfers to a different template or to a different region on the same template. When this transfer is to a different template, the process results in recombination. Retroviruses package two copies of genomic RNA in the virion. These two RNAs may not necessarily be 100% identical. If the two RNAs are not 100% identical, then strand transfers between the two RNAs result in the production of proviral DNA that encodes genomic RNAs that are chimeras of the original parent genomes. Internal strand transfers within the HIV genome have been shown to occur at a high frequency particularly during minus strand DNA synthesis (54, 80, 84, 253, 254). These transfers are distinct from the essential end transfers of – sssDNA and +sssDNA (see Introduction) in that they can occur from any point within the genome. The internal transfer events are proposed to occur by the “forced copy-choice” mechanism (83). The forced copy-choice model proposes that some viral genomic RNAs may be damaged or broken and thereby unable to serve as complete, intact templates to produce a completed copy of minus strand DNA. When DNA synthesis reaches the end of such a broken DNA, the growing DNA is “forced” to transfer to a homologous region on the second RNA template to complete minus strand synthesis. Therefore internal

strand transfers increase the probability of successful DNA synthesis in case of broken or damaged RNA templates by serving as a salvage pathway (83).

Recombination can also occur from intact RNA templates indicating that they are not forced. Hence a slightly modified version of this model termed “copy-choice” is used to describe all recombination events occurring during minus strand DNA synthesis. For example, DNA synthesis along the viral genome can be impeded by certain sequences and/or structures that cause synthesis to stall. Such positions are referred to as pause sites and may serve the role of break points on intact RNA templates. At a pause site, the RNase H activity of RT takes precedence over the polymerization activity (255) and extensively degrades the RNA template. This degradation destabilizes the donor RNA template-growing DNA hybrid and also clears regions on the DNA that can now bind to the acceptor template. The nascent DNA may dissociate from the donor RNA such that the 3' end region of the primer is free of both the donor and acceptor templates and then transfer to and bind to the acceptor template or the latter may invade and displace the donor RNA (Figure 1-5). In either case, the nascent DNA associates with the acceptor RNA where synthesis continues. NC may serve to reduce pausing by melting out impeding secondary structures thereby facilitating continued synthesis of the nascent DNA on the donor template (256). However, in the presence of acceptor template, NC resolves pause sites with an accompanied increase in strand transfer (257) by enhancing RNase H activity of RT and by promoting strand exchange and annealing of the nascent DNA from donor RNA template to acceptor RNA template. Pausing is not the only driving force for strand transfer events. Low structure can promote recombination on

templates by a mechanism where the acceptor can rapidly associate with the nascent DNA (258). In this mechanism, which has been demonstrated with genome segments from the *env* region, there are no strong pause sites to stall the polymerase and promote strand transfer. However, the low degree of secondary structure of the acceptor template makes it easier for it to rapidly bind to the nascent DNA on the donor. This is probably due to the acceptor not having to be “unwound” before hybridization. Consistent with this, this type of transfer is only modestly stimulated by NC as where transfers in regions with high secondary structure are strongly stimulated (259). Association with the nascent DNA in this case probably occurs several nucleotides 5’ of the 3’ DNA terminus in a region that has been cleared by RNase H activity. After hybridizing the acceptor rapidly “zippers” up the DNA catching up with the 3’ terminus and displacing the donor. Regions with a high degree of homology also allow rapid strand transfers by promoting binding of acceptor RNA to the DNA more easily. Reports have shown that runs of the same nucleotide can serve to promote recombination in the vicinity of the run (224, 260). Strand transfer may also occur by another type of pause-independent mechanism that involves specific structural moieties on the acceptor (193, 261). Nevertheless, strand transfers promote completion of DNA synthesis in case of obstacles encountered during reverse transcription and in the process, serve as a means of generating genetic diversity in the viral population. Jetzt *et al* have shown using single-cycle system that HIV-1 undergoes approximately two to three strand transfer events in a single replication cycle (81) while more recent reports now shown even greater rates: about nine times in T-lymphocytes and thirty times in macrophages in single-cycle replication systems (88).

In this section, experiments conducted showed that strand transfer occurs during the synthesis of long reverse transcription products *in vitro* and is the molecular mechanism involved in synthesis of long DNAs *in vitro*. Another possible mechanism for producing the long products in the aggregates, increased processivity of RT, was ruled out previously (see Chapter 4).

5.2 Materials

Plasmid pBKBH10S was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH, from Dr. John Rossi. This plasmid contains an 8.9 kb SstI fragment (nt 222-9154 of the RNA genome) from HIV-1 BH10 inserted into the SstI site. The fragment has all HIV-1 gene coding regions but does not contain the HIV-1 LTR (210). PCR primers and primers used to prime templates in reverse transcription assays were obtained from Integrated DNA Technologies, Inc. The HIV-RT clone was a generous gift from Dr. Samuel H. Wilson (National Institute of Environmental Health Sciences, Research Triangle Park, NC). HIV-RT was purified according to the protocol described (211). The protein was purified to homogeneity and the purity of the protein was evaluated using Coomassie Blue staining of 10 % SDS-PAGE gels (212). The subunits p51 and p66 of RT were in a 1:1 ratio. Aliquots of HIV RT were stored frozen at -80°C and fresh aliquots were used for each experiment. The HIV NC clone was a generous gift from Dr. Charles McHenry (University of Colorado). NC was purified to apparent homogeneity (as judged from Coomassie Blue staining of 17.5% SDS-PAGE gels (212)) according to the protocol described (123). Quantification was by absorbance at 280 nm using a molar extinction coefficient of 8350 cm⁻¹ M⁻¹ (123). Aliquots of NC were stored frozen at -80 °C, and fresh aliquots were used for each experiment. RNaseH minus (E478>Q) RT was a gift from Dr. Stuart Le Grice, HIV Drug Resistance Program, National Cancer Institute, Frederick, MD. *Taq* polymerase was from Eppendorf. T7 RNA polymerase, SP6 RNA polymerase, DNase I-RNase-free and RNase-DNase-free were from Roche Diagnostics. RNase inhibitor was from Promega. T4 polynucleotide kinase and Restriction enzyme HincII was obtained from New

England Biolabs. Proteinase K was obtained from Eastman Kodak Co. Radiolabeled compounds were obtained from Amersham. Sephadex G-25 spin columns were from Amika Corp. RNA cleanup and Mini-prep kits were from Qiagen. GeneTailor site-directed mutagenesis kit, Topo TA cloning kit and high fidelity Platinum *Taq* DNA polymerase were obtained from Invitrogen. All other chemicals were from Sigma or Fisher Scientific.

5.3 Methods

Site-directed mutagenesis of acceptor RNA templates-The mutant primers listed in Table 5-1 were used to introduce approximately equally spaced mutations into the region from position 1 to 1518 (genome bases 222-1739) of the HIV insert in pBKBH10S plasmid. The mutations correspond to positions- 250, 500, 750, 1000 and 1250 from the 5' end of the HIV insert. The primers carrying each mutation were extended during temperature cycling by high fidelity Platinum *Taq* DNA polymerase obtained from Invitrogen (as per manufacturer's protocol). After temperature cycling, the product with the desired mutation was transformed into MAX Efficiency® DH5 α TM-T1^R One Shot® chemically competent cells. Mini preps were obtained using Qiagen mini-prep kits. Sequencing was done using primer 5'- GTTCTAGGTGATATGGCCTGATG-3' to check for mutation incorporation at positions 250 and 500, primer 5'- GACCAACAAGGTTTCTGTCATC-3' to check for mutation incorporation at positions 750 and 1000 and primer 5'- TCTGGCTGTGTGCCCTTCTTTG-3' to check for mutation incorporation at position 1250.

PCR Amplification of DNA substrates for donor RNA (RNA without dimerization signal and for acceptor RNA)- Two PCR primers, 5'- **GATTTAGGTGACACTATAGGAATTAGATCGATGGGAAAA-3'** and 5'- CTGAAGCTCTCTTCTGGTGG-3' were designed to yield donor RNA templates without the dimer initiation site (bases 18 to 135 of the HIV genome insert in pBKBH10S (239-356 of genomic RNA)) and amplified DNA from position 145 to 1538 (genome bases 366-1759) on the HIV insert. Also, two PCR primers, 5'-

GATTTAGGTGACACTATAGAGCTCTCTCGACGCAGGACT-3' and 5'-
GGCTGTTGGCTCTGGTCTGC-3' were designed to yield RNA templates to be used as
acceptor RNA and amplified DNA from position 1 to 1518 (genome bases 222-1739).
The acceptor RNA lacks the primer binding site present on the above donor. An SP6
promoter sequence (in bold) was included on one primer in each primer pair to allow
transcription of the DNA by SP6 RNA polymerase. PCR reactions were performed with
Taq polymerase according to the enzyme manufacturer's protocol using the provided
buffer. One hundred pmol of each primer was used. Reactions included 30 cycles of
denaturation, annealing and extension at temperatures of 94°C for 1 min, 50°C for 1 min
and 72°C for 2 min, respectively followed by one cycle of extension at 72°C for 5 min.
The PCR products were run on a 1% agarose gel, extracted by dialysis and purified as
described (212), and used to prepare RNA as described below.

Preparation of DNA substrates by asymmetrical PCR- Two PCR primers, 5'-
GATTTAGGTGACACTATAGAGCTCTCTCGACGCAGGACT-3' and 5'-
CTGAAGCTCTCTTCTGGTGG-3' were used at 100 pmol and 1 pmol respectively, in
PCR reactions performed with *Taq* polymerase according to the enzyme manufacturer's
protocol using the provided buffer. Reactions included 50 cycles of denaturation,
annealing and extension at temperatures of 94°C for 1 min, 50°C for 1 min and 72°C for
2 min, respectively followed by one cycle of extension at 72°C for 5 min. The PCR
product which is single-stranded plus-strand DNA was run on a 1% agarose gel, extracted
by dialysis and purified as described (212). The amount of recovered DNA was
determined spectrophotometrically from optical density.

Preparation of RNA substrates- RNAs of approximately 1.9 kb were made by first digesting pBKBH10S with restriction enzyme HincII. The digests were then extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with ethanol. Run-off transcription (performed according to the enzyme manufacturer's protocol) was then conducted using 5 µg of the digested plasmid and T7 RNA polymerase enzyme to generate 1.9 kb RNA. Run-off transcription was also performed using ~5 µg of purified PCR DNAs described above and SP6 RNA polymerase to generate RNAs without the dimerization signal (approximately 1.4 kb) and the acceptor RNA (approximately 1.5 kb). The transcription reactions were treated with 2 µl of 10 units/µl of DNase I-RNase-free enzyme for 15 min to digest away the template DNA. The RNA was purified using the Qiagen RNA cleanup kit. The amount of recovered RNA was determined spectrophotometrically from optical density. The integrity of the RNA was checked on a 1% agarose gel as described before.

RNA-DNA and DNA-DNA Hybridization- DNA primers that bound specifically to the RNA and DNA templates: 5'-CTGAAGCTCTCTTCTGGTGG-3' to the 1.9 kb and 1.4 kb (dimer minus) RNA templates as well as the 1.5 kb DNA template and 5'-GGCTGTTGGCTCTGGTCTGC-3' to the 1.5 kb acceptor template were ³²P-labeled at the 5'-end with T4 polynucleotide kinase according to the manufacturer's protocol. The RNA and DNA templates was hybridized to the complementary labeled primer by mixing primer:transcript at a ~ 1:1 ratio in 50 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol and 80 mM KCl. The mixture was heated to 70°C for 5 min and then slowly cooled to room temperature[1].

Reverse transcription reactions with increasing template and with RNase H minus RT- RNA template-DNA primer hybrids (4 nM final concentration of RNA) were pre-incubated for 5 min along with additional template RNA (12 nM) in the experiment with RNase H minus RT and with increasing amounts (4, 8, 12 and 32 nM) of donor RNA (1.9 kb) in the presence or absence of NC (4 μ M), as indicated, in 21 μ l of buffer (see below) at 37°C. The reactions were initiated by addition of 4 μ l of HIV-RT (80 nM final in reactions) or with 4 μ l of RNase H minus RT (80 nM final in reactions). The following reagents at the indicated final concentrations were also included in the reaction mixtures: 50 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 80 mM KCl, 6 mM MgCl₂, 100 μ M dNTPs, 5 mM AMP (pH 7.0), 25 μ M ZnCl₂ and 0.2 units/ μ l RNase inhibitor. Reactions were allowed to incubate for 75 min. The reactions were stopped by adding 2_[12] μ l of a solution containing 250 mM EDTA (pH 8.0) and 5 ng of RNase-DNase-free enzyme and allowed to digest for 20 min at 37°C. Nine μ l of proteinase K at 2 mg/ml in 1.25 % SDS, 15 mM EDTA (pH 8.0) and 10 mM Tris (pH 8.0) was then added to the above mixture, which was placed at 65°C for 1 hour. Finally 7 μ l of 6X alkaline dye (300 mM NaOH, 6 mM EDTA, 15% glycerol, 0.15% bromophenol blue) was added to the mixture and the samples were resolved on 1% alkaline agarose gel containing 50 mM NaOH and 1 mM EDTA (pH=8). The gels were fixed and dried as described (212). Extended DNA products were observed using a Bio-Rad Molecular Imager FX.

Reverse transcription reactions with DNA templates- DNA template-DNA primer hybrids (4 nM final concentration of RNA) were pre-incubated for 5 min along with additional template DNA (12 nM) in the presence of increasing concentrations of NC

(0.5, 1, 2, 4, 6 and 8 μM) or absence of NC in 21 μl of buffer (see below) at 37°C. The reactions were initiated by addition of 4 μl of HIV-RT (80 nM final in reactions). The following reagents at the indicated final concentrations were also included in the reaction mixtures: 50 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 80 mM KCl, 6 mM MgCl_2 , 100 μM dNTPs, 5 mM AMP (pH 7.0), 25 μM ZnCl_2 and 0.2 units/ μl RNase inhibitor. Reactions were allowed to incubate for 75 min and treated as described above for reverse transcription reactions with RNA template.

Strand transfer time course reaction- RNA template without the dimer signal, (shorter donor) hybridized to DNA primer as described above (4 nM final concentration of RNA in reaction) was pre-incubated for 5 min along with acceptor RNA (12 nM) in the presence of NC (4 μM) in 21 μl of buffer (as in reverse transcription experiment) at 37°C. The reactions were initiated by addition of 4 μl of HIV-RT (80 nM final in reactions). The entire reactions were stopped as described above at time points- 2, 5, 10, 15, 30, 45, 60 and 75 mins and samples were resolved on 1% alkaline agarose gel containing 50 mM NaOH and 1 mM EDTA (pH=8) as described above.

Experiments to determine the rate of strand transfer during synthesis of long DNA products in vitro- The 1.9 kb RNA template (donor) was hybridized to DNA primer as described above (4 nM final concentration of RNA in reaction) and pre-incubated for 5 min along with acceptor RNA (4 nM in one reaction and 16 nM in another reaction) in the presence of NC (4 μM) in 21 μl of buffer (as in reverse transcription experiment) at 37°C. Control reactions (donor and acceptor) were also set up. In the donor control

reaction, the 1.9 kb RNA template (donor) was hybridized to DNA primer as described above (4 nM final concentration of RNA in reaction) and pre-incubated for 5 min along with excess of same donor RNA (12 nM). In the acceptor control reaction, the 1.5 kb acceptor RNA template (having the five mutations) was hybridized to DNA primer as described above (4 nM final concentration of RNA in reaction) was pre-incubated for 5 min along with excess of same acceptor RNA (12 nM) in the presence of NC (4 μ M) in 21 μ l of buffer (as in reverse transcription experiment) at 37°C. Assay conditions were as described above except that 50 μ l reactions were performed. The reactions were initiated by addition of 8 μ l of HIV-RT (80 nM final in reactions) and stopped as described above after 75 mins. Reactions were processed, and then electrophoresed on 5 % polyacrylamide denaturing gels. Full length DNA products were located by autoradiography, excised, and eluted overnight in a TE buffer (10 mM Tris-HCL, pH 8.0, 1 mM EDTA, pH 8.0). The eluate was separated from the gel by centrifugation and subsequent filtration through a 0.45-micron disposable syringe filter. The DNAs were recovered by precipitation in ethanol with 300 mM sodium acetate. The recovered DNA was amplified by PCR. Approximately equal amounts (0.2 fmoles each) (as judged by counts per minute (cpm)) of the samples detected using a scintillation counter) of full-length DNA from donor control and acceptor control reactions were mixed and amplified by PCR. The PCR control reaction was set up to check if recombination occurs during PCR amplification of the long DNA products. Primers, 5'-GATTTAGGTGACACTATAGAGCTCTCTCGACGCAGGACT-3' and 5'-CTGAAGCTCTTCTGGTGG-3' were used to PCR amplify the test samples and primers, 5'-GATTTAGGTGACACTATAGAGCTCTCTCGACGCAGGACT-3' and 5'-

GGCTGTTGGCTCTGGTCTGC-3' were used to amplify the PCR control reaction. PCR reactions were performed with *Taq* polymerase according to the enzyme manufacturer's protocol using the provided buffer. Thirty-two pmol of each primer was used. Reactions included 25 cycles of denaturation, annealing and extension at temperatures of 94°C for 1 min, 50°C for 1 min and 72°C for 3 min, respectively followed by one cycle of extension at 72°C for 5 min. The PCR products were run on a 1% agarose gel, extracted by dialysis and purified as described (212). The purified PCR products were ligated into Topo vector (Invitrogen), which was used to transform Top10_ *E. coli* competent cells (as per manufacturer's protocol). Only white colonies were picked. Minipreps were prepared using a Qiagen miniprep kit, and each clone was sequenced using three primers- M13 reverse primer, T7 promoter primer and 5'-GACCAACAAGGTTTCTGTCATC-3'.

Gel electrophoresis- One percent alkaline agarose gels containing 50 mM NaOH and 1 mM EDTA (pH=8), 1 % native agarose gels in Tris-Borate-EDTA buffer and denaturing 5% polyacrylamide gels (19:1) (acrylamide:bisacrylamide), containing 7 M urea were prepared and subjected to electrophoresis as described (212).

5.4 Results

Reverse transcription using RNase H minus RT enzyme (E478>Q) does not produce full-length DNA products- Strand transfer is known to require RNase H activity (262-265). Therefore, if the long DNAs observed in NC reactions were produced by strand transfer then production should be sensitive to RNase H. To test this reactions were performed with an RNase H minus form of HIV-RT (E478>Q) that had wild type polymerase activity (95, 262-265) (Figure 5-1). In the absence of NC this enzyme yielded products that were on average longer than wild type although the total synthesis was the same as wild-type. A very small proportion of fully extended products were also observed with the mutant. In contrast to wild type RT, in the presence of NC E478>Q did not produce any long fully extended DNAs. Only a small amount of short extension products were observed. The sensitivity to RNase H activity is consistent with a strand transfer mechanism being required for long product production (see Discussion).

Full-length DNA is not synthesized when DNA is used as template- Copy-choice type strand transfer does not occur efficiently on DNA templates (225). Therefore a strand transfer mechanism that produces long DNAs with RT and NC should not function efficiently on a DNA template. To test this, a reverse transcription assay was performed using DNA as template. Increasing concentrations of NC (0, 0.5, 1, 2, 4, 6 and 8 μM) and 80 nM RT were included in the reactions which were stopped after 75 min (Figure 5-2). Many of the products produced on the DNA were nearly full length even in the absence of NC, although full length (1538 nucleotides) products were not observed. The average length of extension products in the absence of NC was clearly greater than was observed

with RNA (see Fig. 5-2 for example). The greater efficiency was probably due to the DNA not being susceptible to RNase H activity. This activity can cause the nascent DNA and RNA template to dissociate at times, making extension more difficult on RNA. The major effect of NC on the reactions was to inhibit extension as the level of extended products clearly decreased with increasing NC. This was also observed with RNA (see Chapter 2) along with a dramatic increase in the proportion of full-length products. No increase in longer products was observed with the DNA template. The average length of extended products actually showed some decrease in the presence of NC. Overall the results show that the mechanism responsible for producing long products with the RNA template does not function on DNA. This again supports a strand transfer model.

Synthesis of long DNA products involves strand transfer-The synthesis of long DNA products required RNase H activity of RT and could not be mimicked using a DNA template suggesting that strand transfer is required for long product formation. In order to determine this, a strand transfer time course assay was done that simulates strand transfer events occurring during minus strand synthesis. Donor RNA (dimer minus RNA, 1394 bases long), the template on which DNA synthesis initiated and 3-fold excess acceptor RNA (1518 bases long and containing 5 mutations, see Methods), the template to which DNAs initiating on the donor can potentially transfer, were mixed in reactions. DNA synthesis was initiated from a 5' end-labeled DNA primer that is designed to bind only to the 3' end of the donor RNA (see Figure 5-3, A). Strand

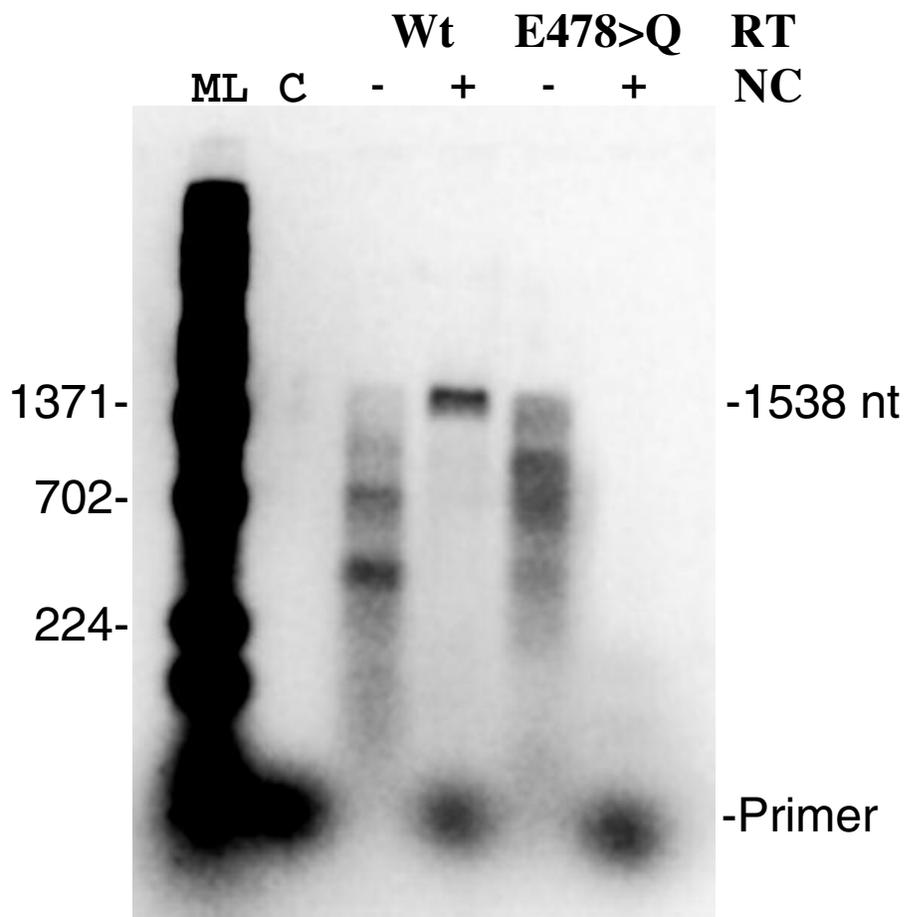


Figure 5-1: Reverse transcription assay using RNase H minus RT enzyme (E478>Q).

Shown is an autoradiogram of an assay in the presence and absence of NC using RNase H minus RT that has wild type polymerase activity with 1.9 kb template as substrate. Lane C is the control reaction in the absence of RT. Lanes with – indicate reactions without NC and those with + indicate reactions with NC (4 μ M). Wt and E478>Q denote reactions with wild type RT and RNaseH minus RT respectively. ML denotes lane with molecular marker (in nucleotides).

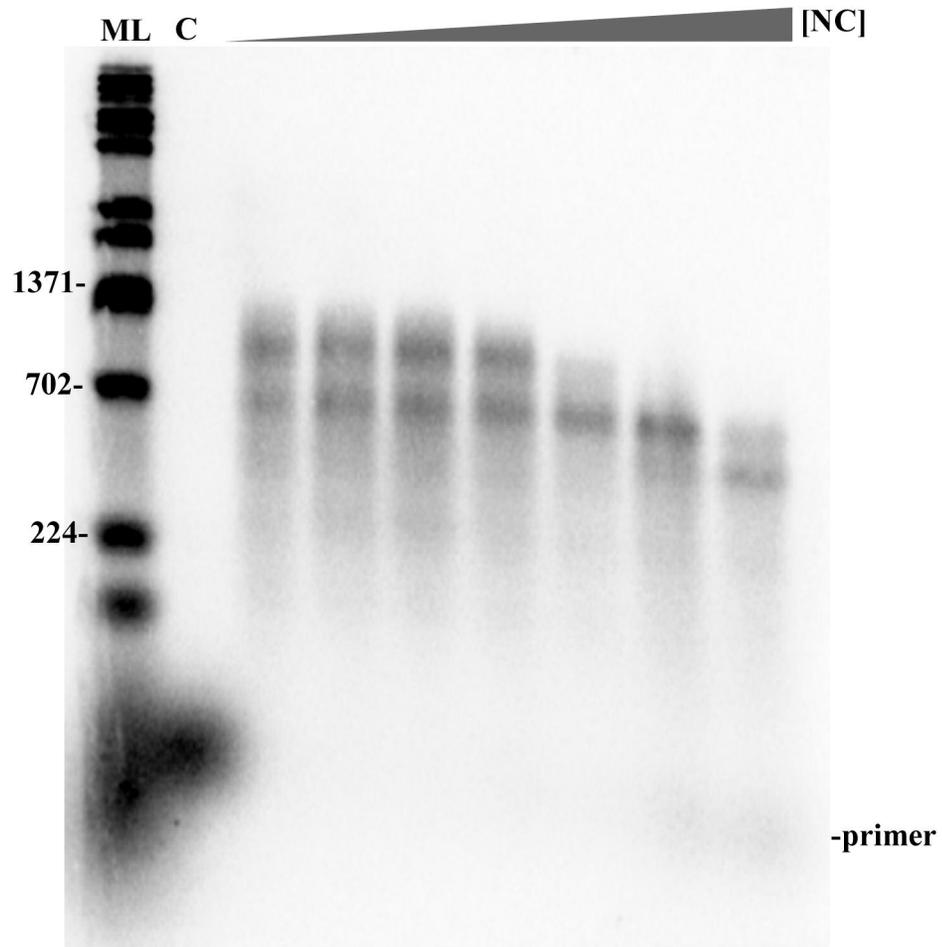


Figure 5-2: Reverse transcription assay with DNA as template.

Shown is an autoradiogram of an assay using 1.5 kb DNA as template in the presence of increasing concentrations of NC (0, 0.5, 1, 2, 4, 6 and 8 μM) and 80 nM RT. C denotes the lane with reaction carried out without RT. ML denotes lane with molecular marker (in nucleotides).

transfer can occur at any point after primer extension occurs on the donor since the donor and acceptor are homologous over this region except for the five mutations described above. The acceptor contains 144 additional bases at the 5' end such that products transferring to and subsequently extended on the acceptor will be 1538 nucleotides as compared to 1394 for extension on the donor. Shown in Figure 5-3, B, is an autoradiogram of a strand transfer time course assay in the presence of 4 μ M NC. A reaction in which excess donor template rather than acceptor was added is also shown to mark the position of products made on the donor (far right lane). Transfer products, migrating slightly higher than full length donor-directed products, were observed in the reactions by 45 min and increased up till 60 minutes. In reactions with acceptor, no products consistent with full-length donor-directed products were observed at any time point. This suggests that transfer to the acceptor occurs before the end of the donor is reached (internal strand transfer). A time course reaction in which the excess acceptor was replaced by excess donor showed essentially the same time frame for appearance of full-length products as shown in Figure 5-4. The results indicate that the long DNAs are produced by strand transfer. The fact that no fully extended donor-directed DNAs were observed indicates that transfers occur before the end of the donor template is reached (see Chapter 6 also).

Determination of the rate of strand transfer- In order to determine that rate of strand transfer during the synthesis of long DNA products an acceptor template with several mutations was made. The acceptor in this case was the same as the donor except it lacked the 20 base primer binding site and contained 5 equally spaced nucleotide

changes at 250 nucleotides intervals. Donor RNA was prehybridized to primer at a 1:1 ratio with the primer. Four-fold excess acceptor RNA (16 nM acceptor, 4 nM donor) was added in reactions with 4 μ M NC and 80 nM RT. Long DNA products were isolated and amplified by PCR, cloned and sequenced as described in Methods. Transfer between each of the mutations could be easily monitored because the products would contain the bases from the donor until transfer to the acceptor took place, at which point a base change to one of the 5 different bases would occur in the sequence. Therefore, there were five 250 base regions within the donor where transfers could be detected (see Figure 5-5). The actual cross-over could have occurred anywhere in the 250 base region, therefore, an assumption was made that each jump occurs on average at the 125th base in between two adjacent mutations. The position of the first transfer from any given donor was calculated by adding 250 for each interval before the interval where cross-over occurred then adding an additional 125 for the cross-over interval. For example, if cross-over occurred between the 2nd and 3rd mutation, the recombination position for the DNA would be 250 + 250 + 125 or 625. The average distance before the first transfer was calculated by summing the numbers from each clone and dividing by the total number of sequence clones. Those clones that had DNAs which jumped to the acceptor before the first mutation were assigned the score of 125. There were six such clones. One clone had DNA that jumped before the fifth mutation and was assigned the score of 1125. One clone had DNA that jumped before the 2nd mutation and was assigned the score of 375. From these, the frequency of strand transfer was calculated as $((6 \times 125) + 1125 + 375) / 8 = 281$. If this number is used to calculate the overall rate of transfer then an estimate of about 1 jump per 300 bases is reasonable. This may not be completely accurate as the other donors

RNAs in the reactions may also serve as acceptors and donor to donor jumps cannot be detected. The fact that six of the eight clones that were sequenced had jumped back to donor after going to the acceptor clearly indicates that donors can serve as “acceptor” in the reactions. Jumps to the donor should be limited because of the 4 fold excess of acceptor in the reactions; however, the effective concentration of the acceptor and donor may be slightly different than the actual concentration ratio of 4:1. Overall, if some of the first jumps were to other donors these would not have been observed and the actual transfer rate may be slightly greater than one per 300.

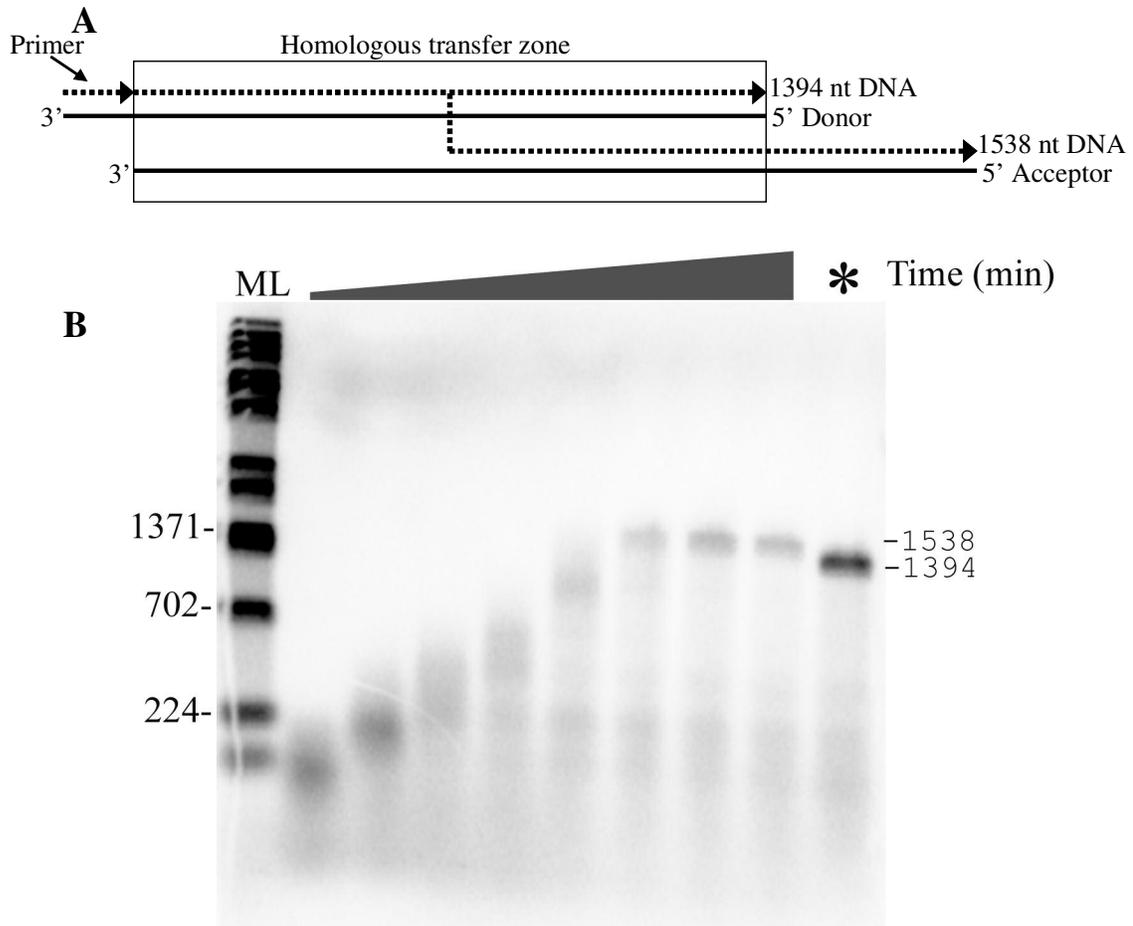


Figure 5-3: Time course strand transfer assay shows that strand transfer is involved in the synthesis of long DNA products.

(A) A schematic diagram of the strand transfer assay is shown. The RNA (solid lines) without the dimer signal was used as the donor and was hybridized to a 20 nucleotide 5' P-32 labeled DNA (dashed lines) primer. A longer RNA without the binding site for the donor primer was used as acceptor. Full extension on the donor produced a 1394 nucleotide DNA while transfer and subsequent extension on the acceptor produced a 1538 nucleotides product. The region of homology (“homologous transfer zone”) between the donor and acceptor is boxed. (B) An autoradiogram of a strand transfer experiment. The reactions were carried out in presence of 4 μ M NC for time points 2, 5, 10, 15, 30, 45, 60 and 75 mins as shown from left to right. In the last lane (*), the reaction was carried out by adding excess donor RNA to the reaction instead of acceptor RNA and this reaction was carried out for 75 min. The positions of full length DNA products from the donor (1394) or transfer products (1538) are indicated. ML denotes lane with molecular marker (in nucleotides).

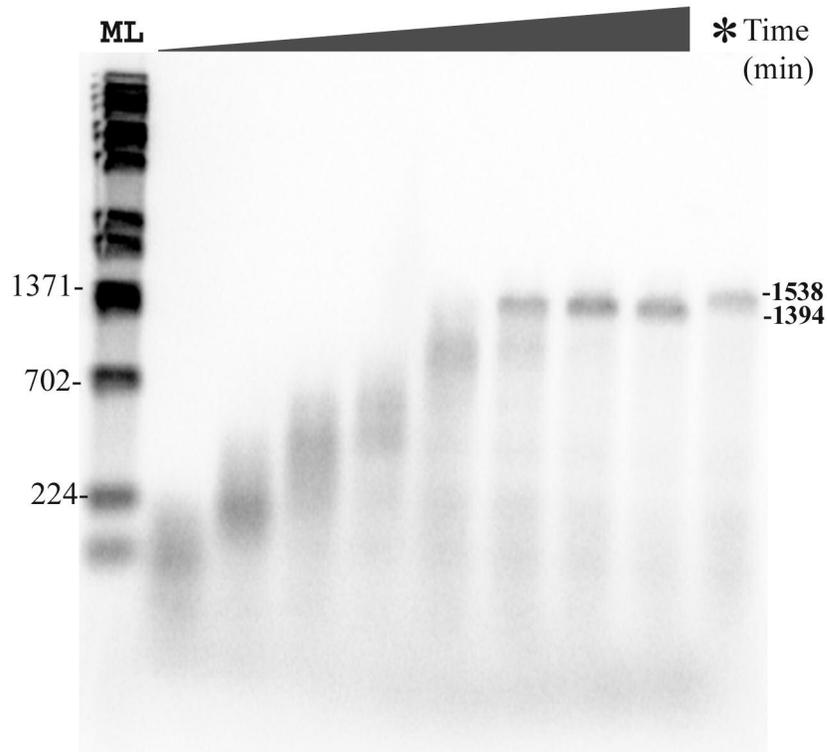


Figure 5-4: An autoradiogram of a time course reaction in which excess acceptor was replaced by excess donor.

The reactions were carried out in presence of 4 μ M NC for time points 2, 5, 10, 15, 30, 45, 60 and 75 mins as shown from left to right. In the last lane (*), the reaction was carried out by adding excess acceptor RNA to the reaction instead of donor RNA and this reaction was carried out for 75 min. The positions of DNA products from the donor (1394) or acceptor (1538) are indicated. ML denotes lane with molecular marker (in nucleotides).

Table 5-1: Primers for synthesis of mutations in acceptor

Base ^a	position on HIV genome ^b	Sequence of primers in the 5' and 3' orientations ^c
250	471	5'-gagctagaacgattcgcagtt <u>t</u> atcctggcctgtag-3' 5'-aactgcgaatcgttctagctccctgctgcccatac-3'
500	721	5'-gacacagcagtcaggtcagcc <u>t</u> aaattaccctatag-3' 3'-ggctgacctgactgctgtgctctgtgcagc-3'
750	971	5'-gaggaagctgcagaatggga <u>a</u> agagtacatccagtg-3' 5'-tcccattctgcagcttctcattgatggctc-3'
1000	1221	5'-aagaaccttttagagactat <u>t</u> tagaccggtctataaac-3' 5'-atagtctctaaaaggttcttttggctcctgte 3'
1250	1471	5'-cagctaccataatgatgcaga <u>c</u> aggcaattttaggaac-3' 5'-tctgcatcattatggtagctgtattgttacttg-3'

^aThe base number refers to the sequence number of the provirus in pBKBH10S plasmid.

^bThis number refers to the position of the nucleotide on the HIV genome.

^cPrimer pairs that are complementary to opposite strands of pBKBH10S plasmid. One of the primers in each primer pair has the desired mutation indicated by underlined bolded letters.

No of clones sequenced	Base position of mutation				
	1250 (5 th)	1000 (4 th)	750 (3 rd)	500 (2 nd)	250 (1 st)
1	A	A	A	A	A
2	A	D	D	D	D
3	A	D	D	A	D
4	A	A	A	A	D
5	D	D	D	D	A
6	D	A	A	A	D
7	A	A	A	A	D
8	A	A	A	A	A

Table 5-2:

This table is a grid which shows the nucleotide copied from donor or acceptor at the 5 mutated base positions for eight clones that were sequenced. D indicates that the base at that position was copied from the donor and A indicates that the base at that position was copied from the acceptor. Cross-over to the acceptor occurs before the position marked in red.

5.5 Discussion

The formation of high molecular weight aggregates (discussed in Chapter 3) allows the RNA molecules to be in close proximity to each other thereby allowing strand transfers to occur more easily. This could further be facilitated by the high local concentrations of NC and RT in the complexes. Tanchou *et al* (236) have shown that strand transfer occurs within HIV-1 nucleoprotein complexes *in vitro*. Experiments using RNase H minus mutant HIV-RT (E478>Q) that lacks RNase H activity but retains full polymerase activity showed that synthesis of long DNA products is dependent on RNase H activity (Figure 5-1). It is well known that the RNase H activity of RT is required for strand transfer (95, 262-264). This suggests that full-length DNA products result from strand transfer, especially since no increase in processivity was observed in the reactions (Chapter 4, Figure 4-2). Furthermore, time course experiments using donor and acceptor RNAs showed that the long DNA products resulted from internal strand transfer events (Figure 5-3). All the products synthesized transferred to the acceptor since there were no donor directed products in the reaction. This demonstrates that synthesis of full-length DNA products involves the mechanism of internal strand transfer.

Although in the absence of NC products with DNA templates were on average longer than those with RNA, no full-length DNA products were formed at any concentration of NC when DNA was used as template. This is typically what is seen in the cell. Ninety-nine % of the time, the plus strand DNA is not completed and these incomplete and discontinuous plus strands may or may not get integrated into the host chromosome (266). The incomplete strands that get integrated are later on completed by

host polymerases and ligase. Other retroviruses like AMV are known to synthesize plus strands discontinuously with the completed strand being formed by ligation of several smaller products in the nucleus. Unlike the plus strand DNA, the minus strand must be completed continuously for successful replication to occur, because without a complete minus strand the plus strand cannot be made. It is therefore not surprising that retroviruses would have evolved a transfer mechanism to ensure production of intact minus strands.

The rate of strand transfer observed in this system was approximately one jump for every 281 nucleotides. As was noted in the Results, the rate could be somewhat lower but an estimate of 1 per 250-300 nucleotides is reasonable. One thing that could have influenced the rate is a recombination “hotspot” in the first interval. If the nucleotide sequence or structure in the first 250 base interval was more conducive to transfer than other intervals this could have led to more transfers in the first zone. In hindsight, having more than 5 mutations in the acceptor would have made it easier to make a more accurate calculation. Since more than half of the clones crossed to the acceptor before the first mutation, it was clear that transfer was very frequent but this only indicated that it was at least one transfer per 250 bases. More mutations, especially in the first interval could have further defined the rate. The danger of adding more mutations is that this would affect homology and therefore could influence the recombination rate. As will be seen in Chapter 6, homology is a very important driving force for recombination and even clones that differ by a small amount can show highly different transfer rates.

With respect to the recombination rates estimated in cell infections (see Introduction to this chapter), the rates in the *in vitro* system were comparable to rates calculated in macrophages (about 30 recombination over the approximately 9,000 base genome or one per 300 bases) but considerably higher than those in T cells (about 1 in 1,000). Rates *in vitro* may be greater because each aggregate probably contains several hundred or more templates (based on the aggregates centrifuging in microfuge in less than 1 minute) where viruses only have two genome copies.

Chapter 6 Characterization of ‘hotspots’ for strand transfer in the C3 *env* region during Human Immunodeficiency Virus-Type 1 (HIV-1) intersubtype recombination

6.1 Introduction

Human Immunodeficiency Virus type-1 (HIV-1) is known to have very high evolutionary potential in the human host. This is due to its high mutation rate (3×10^5 mutations/site/generation) (267) and recombination during replication (268), extensive replication (10^8 to 10^9 virions per day) (269, 270) and large numbers of infected cells (271). Most of these mutations are detrimental to the virus and are eliminated by negative selection (272). The rate at which mutants accumulate in a population is determined by the error rate of replication as well as by the fitness of the mutants that appear. Any mutation that increases the fitness of the virus will tend to increase in the population whereas the frequency of mutations that decrease the fitness of the virus will remain low. Selective forces that increase survival of particular mutants include the host humoral and cellular immune responses, adaptation of the virus in a new host having a different genetic background from the earlier one, adaptation of the virus to a particular tissue or cell type that it may have invaded as well as any antiretroviral therapy administered to treat the host. HIV isolates that emerge in the human host as a consequence of selective pressure and genetic variability often show increased replicative and cytopathic capabilities. HIV-1 populations therefore exist within their hosts as ‘quasispecies’ (79, 271). This increase in genetic diversity enables the virus to escape the humoral (273, 274)

and cytotoxic T-lymphocytes (CTL) (275-277) immune responses of the host as well as develop resistance to available antiretroviral therapy (278).

Genetic changes are observed throughout the genome, however variation in *env* accounts for much of the differences among the isolates (279). In the *env* region, nucleotide substitutions accumulate at an average rate of about 2.5% a_[J3] year and amino acid substitutions occur at an average rate of about 1% per year in an individual. Specific hyper variable regions exist within the *env* gene, interspersed among more conserved regions (280, 281). Within these hyper variable regions, a remarkably high percentage (>95%) of the nucleotide substitutions found in isolates result in change of an amino acid_[J4]. Hyper variable regions are not likely sites of frequent errors but are rather likely to be regions where a genetic change can be tolerated or where mutations confer selective advantage to the virus (282). These are most likely antigenic variants selected by host immune response and/or variants selected for cell tropism. One of the major goals of HIV research is to determine the selective forces that drive genetic variation in each of the variable regions of *env* gene.

Recombination between variants of HIV also contributes to genetic changes within human hosts. Recombination occurs between HIV progeny derived from cells infected with two different viruses and carrying a genome from each (called heterodiploid viruses). This implies that recombination is likely to be limited by the frequency of doubly infected cells. Viral genomes containing recombinant sequences of HIVs belonging to different subtypes within the HIV major group (A-K) have been isolated

from patients, indicating a high frequency of co-infection of individuals with different viruses (at approximately the same time) and sufficient co-infection of cells within an infected individual to allow recombination to occur during replication (283). Interestingly, although co-infection of the same cell by two different viruses would be predicted to be low, specific mechanisms favoring co-infection have recently been uncovered (284, 285). This may help explain the high level of intersubtype recombinants found worldwide. In Thailand for example, the predominant HIV form is an A/E recombinant (CRF01_A/E) while A/D recombinants are highly prevalent in Uganda (286, 287). The rise of intersubtype recombinants further complicates efforts to produce vaccines against HIV and diminishes the possibility of producing one vaccine with high efficacy against all strains of HIV.

In East Africa and particularly in Uganda, subtypes A and D of Group M of HIV-1 are found to co-circulate with a high prevalence (50% subtype A, 40% subtype D) (287, 288). Co-circulation of sub-types of HIV among individuals often gives rise to unique recombinant forms (URFs) eg. A/D recombinants found in Uganda. Continuous human to human transmission of recombinant forms having well defined chimeric genomes gives rise to circulating recombinant forms (CRFs). A/D URFs are predominant in Uganda as opposed to the more stable CRFs (287, 288).

In an effort to better understand how these recombinants arise, a collaboration between our group and those of Drs. Eric Arts (Case Western Reserve University) and Matteo Negroni (Institut Pasteur) has been established. Dr. Arts has access to patient

isolated from Uganda while Dr. Negroni is an expert on retrovirus recombination and has developed an assay to test the level and position of recombination in a single cycle of infection (289, 290). This assay essentially tests for “hotspots” within the genome where recombination occurs with higher frequency. The single cycle nature of the assay eliminates selection so that recombinants leading to viable or non-viable virus can be examined. Dr. Arts has developed assays that can examine recombinants during several rounds of replication in cells such that the fate of various recombinants can be determined in the context of selection for infectivity in cells. Using these approaches intersubtype recombination between several HIV A and D subtype patient isolates from Uganda was studied. Intersubtype recombination frequencies within the *env* gene increased with an increase in sequence identity between the isolates and hotspots were observed only in the conserved regions (denoted C1-C5) of the *env* gene where sequence homology was higher (Baird et al. 2006 submitted). An interesting finding was a particular A subtype (denoted A115) that showed a hotspot in C3. This subtype was the only clone to show a C3 hotspot. Our lab has expertise in uncovering mechanisms of recombination. This section describes preliminary experiments done in our lab to determine why A115 but not other clones show a hotspot for recombination within the C3 region.

6.2 Materials

Sub-clones A120, A115 and D89 containing HIV-1 envelope gene fragments from subtypes A120, A115 and D89 (HXB2 nt 6420-7520) respectively, were a gift from Dr. Eric Arts (Case Western Reserve University, Cleveland, Ohio). PCR primers and primers used to prime templates in reverse transcription assays were obtained from Integrated DNA Technologies, Inc. The HIV-RT clone was a generous gift from Dr. Samuel H. Wilson (National Institute of Environmental Health Sciences, Research Triangle Park, NC). HIV-RT was purified according to the protocol described (211). The protein was purified to homogeneity and the purity of the protein was evaluated using Coomassie Blue staining of 10 % SDS-PAGE gels (212). The subunits p51 and p66 of RT were in a 1:1 ratio. Aliquots of HIV RT were stored frozen at -80°C and fresh aliquots were used for each experiment. The HIV NC clone was a generous gift from Dr. Charles McHenry (University of Colorado). NC was purified to apparent homogeneity (as judged from Coomassie Blue staining of 17.5% SDS-PAGE gels (212)) according to the protocol described (123). Quantification was by absorbance at 280 nm using a molar extinction coefficient of 8350 cm⁻¹ M⁻¹ (123). Aliquots of NC were stored frozen at -80 °C, and fresh aliquots were used for each experiment. *Taq* polymerase was from Eppendorf. SP6 RNA polymerase, DNase I-RNase-free and RNase-DNase-free were from Roche Diagnostics. RNase inhibitor was from Promega. T4 polynucleotide kinase was obtained from New England Biolabs. Proteinase K was obtained from Eastman Kodak Co. Radiolabeled compounds were obtained from Amersham. Sephadex G-25 spin columns were from Amika Corp. RNA cleanup and mini-prep kits were from Qiagen. All other chemicals were from Sigma or Fisher Scientific.

6.3 Methods

PCR Amplification of HIV-1 envelope gene fragments- Primer pairs (5' **GATTTAGGTGACACTATAGATATAATGAGGTAGTCAAACAATTA-3'** and 5'-TTTATTCTGCATTTGAGAGT-3' for A115 Donor 5'-**GATTTAGGTGACACTATAGATATAAGGAGGTAGCCAAACAATTA-3'** and 5'-TTTATTCTGCATTGGAGAGT-3' for A120 Donor, 5'-**GATTTAGGTGACACTATAGTATATAGAATGGAATAAACTATAC-3'** and 5'-ACCGTTTGTGTTTGTACTCT-3' for D89 Acceptor and 5'-**GATTTAGGTGACACTATAGTATATAGAATGGAATATAGCTTTGT-3'** and 5'-TAGATTGGCATTGACATAG-3' for A115 Acceptor were used in PCR reactions to amplify nts 7255-7474 (relative to HXB-2 provirus numbering) for 115-A and 120-A, and 7235-7454 for 89-D. The bolded regions of the primers are SP6 promoter sequences while italicized regions are non-HIV sequences. PCR reactions were performed with *Taq* polymerase according to the enzyme manufacturer's protocol using the provided buffer. One hundred pmol of each primer was used. Reactions included 30 cycles of denaturation, annealing and extension at temperatures of 94°C, 50°C and 72°C respectively for 1 min, followed by one cycle of extension at 72°C for 5 min. The PCR products were run on an 8% native polyacrylamide gels and purified as described (212).

Preparation of RNA substrates- Run-off transcription (performed according to the enzyme manufacturer's protocol) was conducted using approximately 5 µg of each of the amplified DNAs and SP6 RNA polymerase enzyme to generate RNAs of 225 nts. The transcription reactions were treated with 2 µl of 10 units/µl of DNase I-RNase-free

enzyme for 15 min to digest away the template DNA. The RNA was purified using the Qiagen RNA cleanup kit. The amount of recovered RNA was determined spectrophotometrically from optical density.

RNA-DNA Hybridization- A DNA primer that binds specifically to the donor RNA transcript (5'-TTTATTCTGCATTTGAGAGT-3' or 5'-TTTATTCTGCATTGGAGAGT-3' for A-115 and A-120, respectively) was ³²P-labeled at the 5' end with T4 polynucleotide kinase according to the manufacturer's protocol (New England Biolabs). The donor RNA was hybridized to a complementary labeled primer by mixing primer:transcript at approximately 3:1 ratio in 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 80 mM KCl. The mixture was heated to 65°C for 5 min and then slowly cooled to room temperature.

Time course strand transfer assay- Donor RNA-primer DNA hybrids (2 nM final concentration of RNA) were preincubated for 3 min in the presence or absence of 10 nM acceptor RNA template and NC (as indicated) in 42 µl of buffer (see below) at 37°C. One molecule of NC per two nucleotides was used in the reactions_[15] (18 and 4 µM for reactions with or without acceptor, respectively). The reactions were initiated by the addition of 8 µl of HIV-RT (80 nM final in reactions) to a mixture of 50 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 80 mM KCl, 6 mM MgCl₂, 100 µM dNTPs, 5 mM AMP (pH 7.0), 25 µM ZnCl₂ and 0.4 units/µl RNase inhibitor. Reactions were allowed to incubate for 0, 30s, 1, 2, 4, 8, 16, 32, and 64 min at 37°C prior to quenching a 6 µl aliquot of each reaction with 4 µl 25 mM EDTA (pH 8.0) and 2.5 ng of RNase-DNase free

enzyme for 20 min at 37°C. Two µl of proteinase K at 2 mg/ml in 1.25% SDS, 15 mM EDTA (pH 8.0) and 10 mM Tris (pH 8.0) was then added to the above mixture, which was placed at 65°C for 1 hour. Finally, 12 µl of 2X formamide dye (90% formamide, 10 mM EDTA (pH 8.0), 0.1% xylene cyanol, 0.1% bromophenol blue) was added to the mixture and the samples were resolved on an 8% denaturing polyacrylamide gel containing 7 M urea. Extended DNA products were quantified by phosphorimager analysis using a Bio-Rad FX phosphoimager.

Gel electrophoresis- One percent native agarose gels in Tris-Borate-EDTA buffer and denaturing 8% polyacrylamide gels (19:1) (acrylamide:bisacrylamide), containing 7 M urea were prepared and subjected to electrophoresis as described (212).

6.4 Results

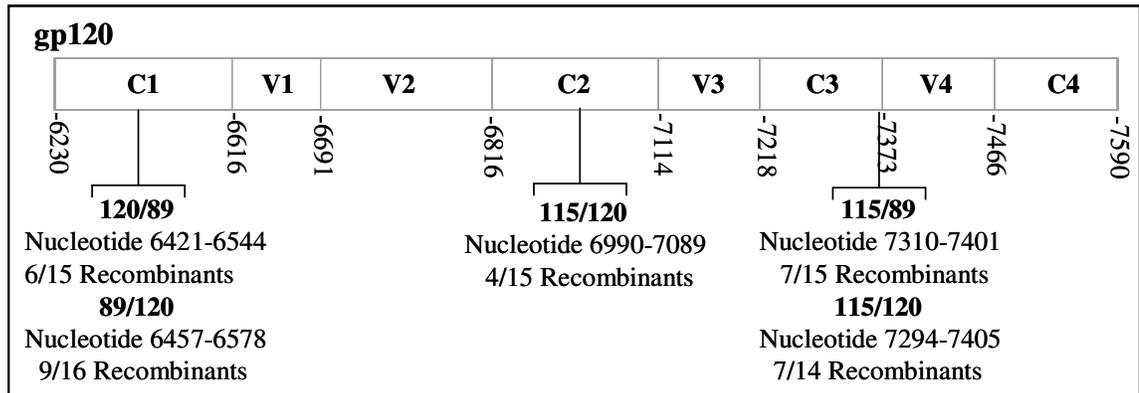
A single cycle cellular recombination assay was used to determine the rate and map the positions of recombination cross-overs between different virus subtypes (Baird et al., 2006, submitted). In this system cells are infected with HIV virus clones that are heterozygous and contain genome segments from two different subtypes. The genome segments are designed such that one serves as a donor and the other acceptor. Sequence analysis of proviruses made during replication allows cross-over points to be mapped. Since the genome segments are from different subtypes natural sequence variation can be used to map the approximate position of cross-overs. Recombination frequencies as measured in the C1-C4 regions of the HIV *env* gene were found to be approximately 1.5 to 2-fold higher when A115 was used as donor than when other subtype A and D genomes segments were used. Analysis of the recombination frequencies and cross-over positions in the C1-C4 *env* region revealed that all intra subtype and inter subtype *env* recombinants had preferential recombination breakpoints (hotspots) in the C1 region (Figure. 6-1A). A hotspot near the V2/C2 junction of *env* was unique to experiments in which D126 was used as the donor. Also, hotspots for recombination were found in the C3 region only in intra and inter subtype *env* recombinants derived from experiments with A115 as the donor.

To further investigate that A115 specific *env* C3 recombination site, the template switching frequency and the RT pausing pattern in the C3 to V4 regions were analyzed using a reconstituted *in vitro* reverse transcription assay. A schematic representation of the assay is shown in Figure 6-1B. Primers annealing to the A115 or A120 donor RNA

templates were radio labeled and hybridized to the respective templates and reverse transcription assay was carried out by adding RT. Template switching to the D89 template during reverse transcription (schematic, Figure 6-1B) was monitored in the presence or absence of HIV-1 NC during a time course assay. A strand transfer time course assay was carried out using A115 as donor and acceptor templates. The addition of non-retroviral nucleotides at the 5' end of the donor template prevents transfer from the end of the donor thereby limiting transfer to the boxed region shown in Figure 6-1B. Analysis of the time course assays shown in Figures 6-2 (A115 donor) and 6-3 (A120 donor) showed that there was more RT pausing on the A115 template as compared to the A120 template during minus strand DNA synthesis, particularly in the C3 region of the RNA template. This pausing was observed irrespective of the presence of acceptor indicating that it originated from DNA synthesis on the donor template. Most of the paused products were chased away at higher time points during the time course reaction. A small percentage of the DNAs transferred to the D89 acceptor template for continued elongation (245 nt product). The frequency of strand transfer events was plotted (Figure 6-5). When A115 RNA was used as donor and acceptor (100% sequence identity in transfer region) in strand transfer assays, the strand transfer efficiency was about 40% without NC and 60% with NC at 64 min (Figure 6-4; 6-5B). However, transfer efficiency was less than 14% with NC with the A115 donor/D89 acceptor pair. This shows that increase recombination during reverse transcription is seen with increased sequence identity. The sequence identities for the A120/D89 pair and A115/D89 were nearly the same at 69% and 66% respectively. However, the transfer efficiency from A115 donor template to D89 acceptor templates was at least 2-fold greater than that observed from

A120 donor template to D89 acceptor template. NC proportionally increased transfer with both template pairs. Therefore, increased strand transfer was observed with A115 donor over A120 donor with or without NC throughout the time course. The results are consistent with the single cycle assays where only A115 donors showed a hotspot in the C3 region. Using primers specific for the strand transfer products, the recombinants from the A115/D89 assays were amplified by PCR and 35 (19 without NC and 16 with NC) clones were sequenced. Figure 6-6, A shows regions of strand transfer from A115 donor to D89 acceptor and Figure 6-6, B shows the number of crossovers within these regions. In the presence of NC, there was a focusing of the transfer points with most occurring in regions denoted VIII-X (11/16). Recombination in this assay matched the C3-V4 recombination regions observed in the single-cycle assay involving the A115/D89 donor acceptor pair. In this case 4 of the 7 total recombinants recovered from the C3 region mapped to positions VIII-X (Baird et al. 2006, submitted).

A_[J6]



B

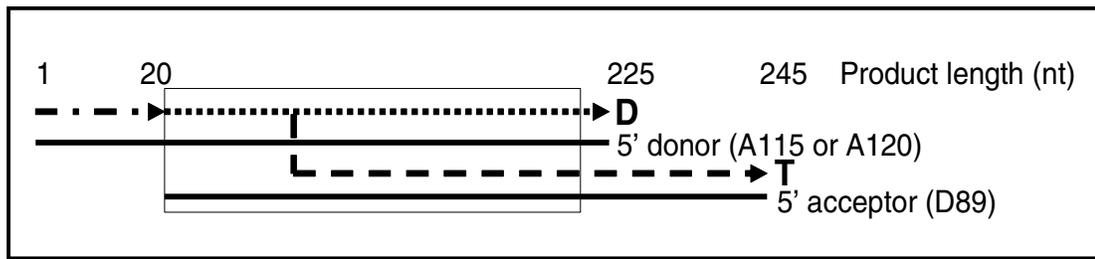


Figure 6-1:

A. In A, is shown the C1-C4 regions of the HIV *env* gene where the recombination frequencies_[J7] were measured in single cycle assays and in multiple round virus replication systems. Recombination hotspots were detected in the C1, C2 and C3-V4 regions of the *env* gene using single cycle and multiple round replication system when the above combinations of subtypes (numbers refer to different HIV subtypes) were used as donor and acceptor. 120/89 means that 120 was used as donor RNA and 89 was used as acceptor template. The C3 to V4 region was analyzed in this *in vitro* reverse transcription assay. **B.** Schematic diagram of *in vitro* template switching assay. The RNA donor template from A115 or A120 was primed with a 20 nt 5' P-32 end-labeled DNA. Extension of the primer by RT to the end of the donor produced a 225 nt product (D) while transfer and subsequent extension to the end of the acceptor RNA (derived from D89) produced a 245 nt product (T). The boxed region is the "homologous" transfer zone (corresponds to bases 7255-7454 of clones based on HIV HXB2 proviral numbering where cross-overs occur).

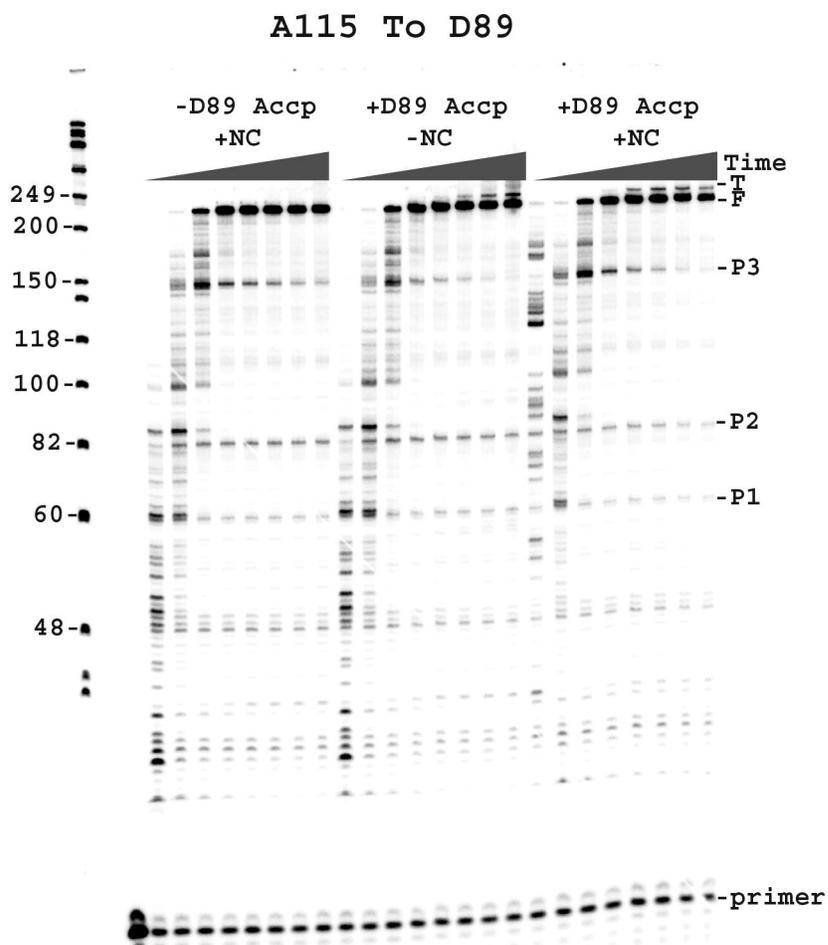


Figure 6-2: Autoradiogram of template-switching assay from A115 to D89.

Shown is an autoradiogram of a template-switching assay from A115 to D89. Positions of primer, transfer (T), and donor-directed (F) products are indicated. The three most prominent pause sites (“P”) are also labeled. Assays were performed in the presence or absence of NC and D89 acceptor as indicated. Times for each set from left to right were: 0.5, 1, 2, 4, 8, 16, 32, and 64 min. A size marker with lengths in nucleotides is shown on the left.

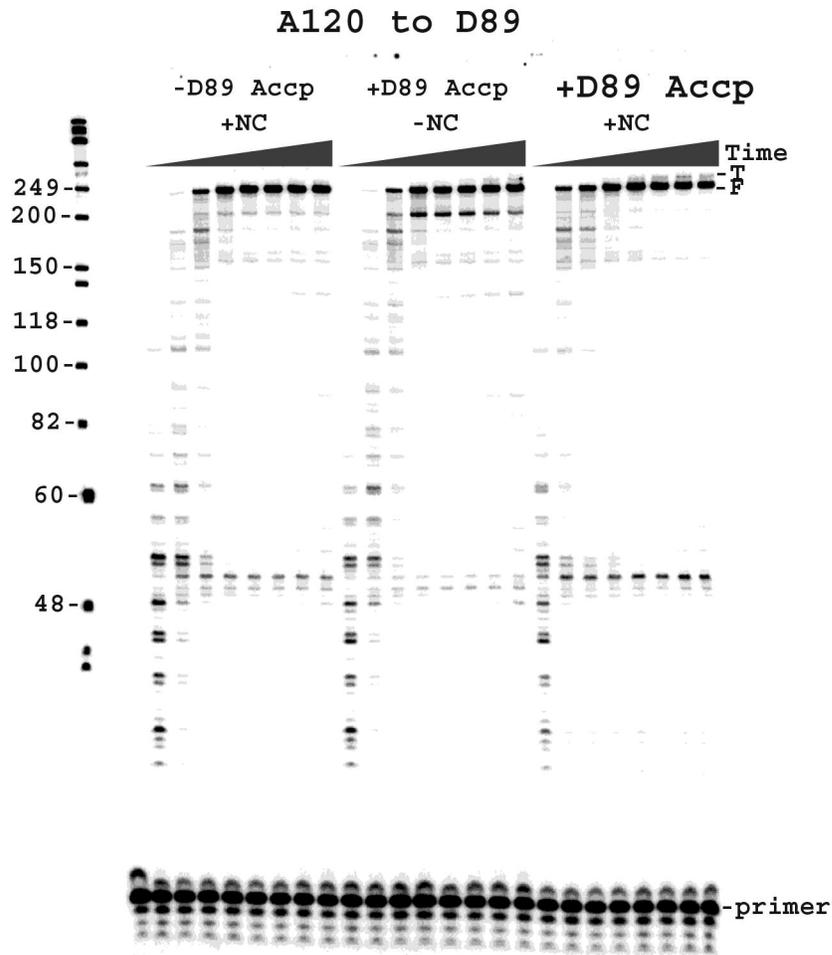


Figure 6-3: Autoradiogram of template-switching assay from A120 to D89.

Shown is an autoradiogram of a template-switching assay from A120 to D89. Positions of primer, transfer (T), and donor-directed (F) products are indicated. Assays were performed in the presence or absence of NC and D89 acceptor as indicated. Times for each set from left to right were: 0.5, 1, 2, 4, 8, 16, 32, and 64 min.

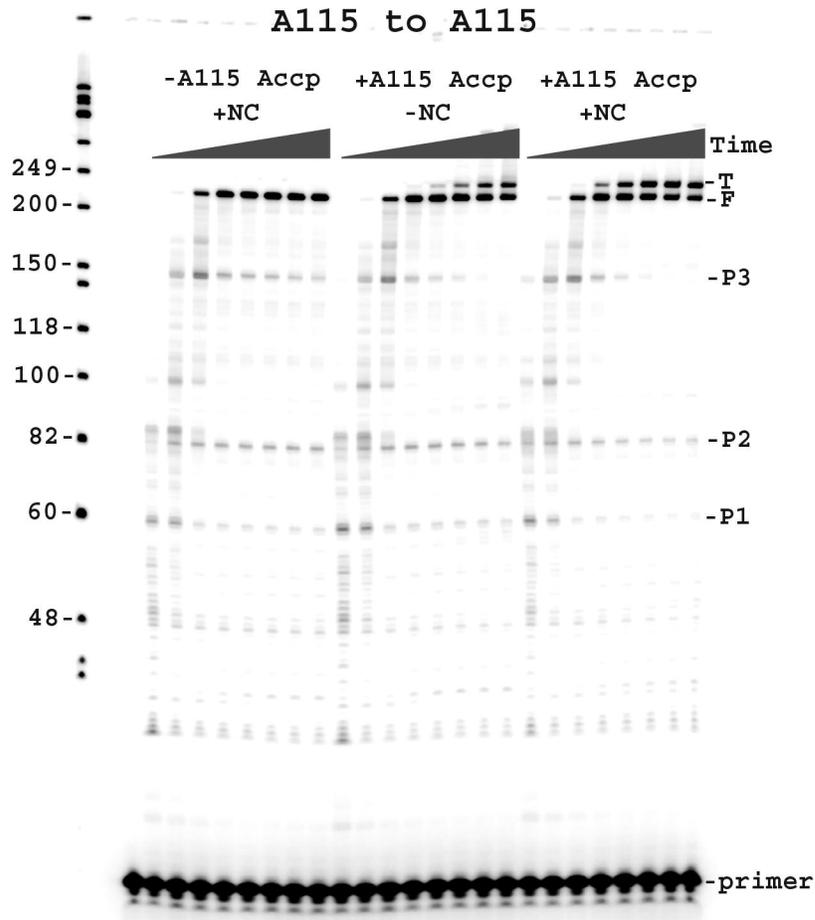


Figure 6-4: Autoradiogram of template-switching assay from A115 to A115.

Shown is an autoradiogram of a template-switching assay from A115 to A115. Positions of primer, transfer (T), and donor-directed (F) products are indicated. The three most prominent pause sites (“P”) are also labeled. Assays were performed in the presence or absence of NC and A115 acceptor as indicated. Times for each set from left to right were: 0.5, 1, 2, 4, 8, 16, 32, and 64 min.

[J8]

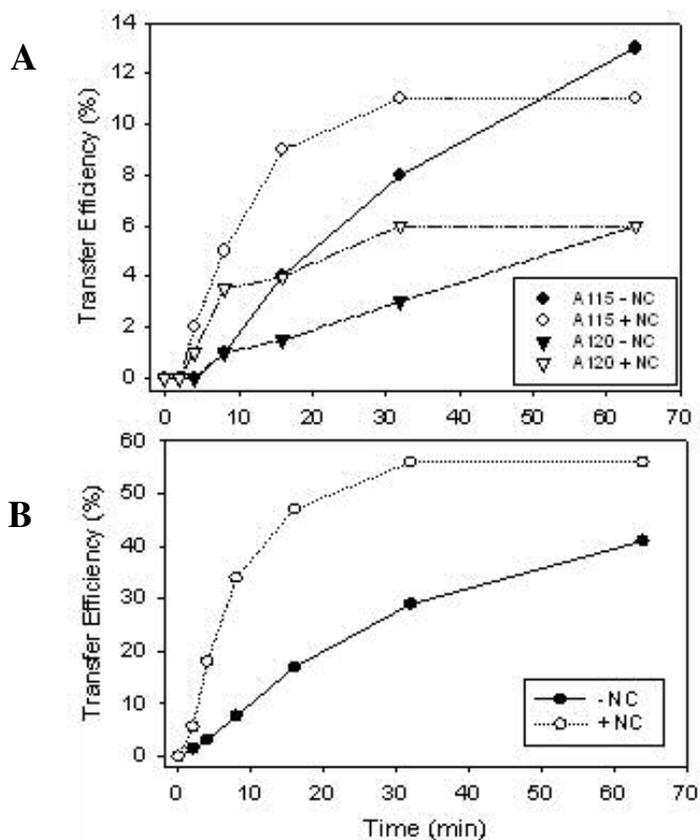


Figure 6-5: Plots of transfer efficiency vs. time

The plot above shows the transfer efficiency vs. time from experiments A115 or A120 to D89 acceptor experiments. Efficiency is defined as: $(\text{transfer}/(\text{transfer} + \text{donor-directed})) \times 100$ or $(T/(T + D)) \times 100$. Shown below is the plot of transfer efficiency vs. time for experiment with A115 as donor and acceptor.

A

D89	7255	-AAC	<u><i>CAGGUAGCU</i></u>	<u><i>AAAAAAUU</i></u>	<u><i>AGGAGCUCUU</i></u>	<u><i>UUUAACA</i></u>	<u><i>AGACAACA</i></u>	<u><i>AAUUUUUCA</i></u>	<u><i>CCA</i></u>	<u><i>UCCUCGG</i></u>	-7321				
A115		UG	UC	C	AACA	AACC	A	AC	A	UAGC	A				
A120		GG	C	C	A	AA	GG	C	AG	C	GUA	GA			
			<u><i>VIII</i></u>	<u><i>P3</i></u>	<u><i>IX</i></u>	<u><i>X</i></u>	<u><i>XI</i></u>	<u><i>XII</i></u>	<u><i>XIII</i></u>	<u><i>XIV</i></u>					
7322	-	<u><i>AGGGG</i></u>	<u><i>ACC</i></u>	<u><i>UAG</i></u>	<u><i>AAUU</i></u>	<u><i>ACAACA</i></u>	<u><i>CAGC</i></u>	<u><i>UUUAAUUGUGGAGG</i></u>	<u><i>GGAAUUUUUC</i></u>	<u><i>UACUGCAAU</i></u>	<u><i>ACAUCA</i></u>	-7388			
		U	A	U	U	A	A	G	U	U	C				
		U	A	U	UG	A	G	U	U	C					
		<u><i>P2</i></u>		<u><i>P1</i></u>											
7389	-	AGACUGUUUAAU	AGGACGUGG	ACGUC	AAAUAGUACA	AUUGGUA	AAUUGAGAGUACA	AAACACAA	ACGGU	-7454					
		G	C	C	AC	UU	C	C	GUG	AACG	CACU	UGU	UG	C	UCUA
		C	C	A	GAUAAUGCC	CUUGCA	GGUCA	UGAC	G	GU	G	C			

Transfer Region	(-)NC	(+)NC
B I	1/19	-
II	1/19	-
III	1/19	-
IV	4/19	-
V	1/19	1/16
VI	1/19	-
VII	1/19	1/16
VIII	3/19	4/16
IX	4/19	3/16
X	-	4/16
XI	-	1/16
XII	1/19	-
XIII	-	1/16
XIV	1/19	1/16

Figure 6-6: (A) Transfer map showing regions of strand transfer from A115 donor to D89 acceptor. The regions are denoted with roman numerals (I-XIV), underlined, and alternately shown in black and gray italics. Residues of A115 (and A120) donor differing from D89 are indicated. Cross-overs were not recovered from regions that are not underlined. Numbering is based on the HIV HXB2 provirus. Nucleotides 7255-7373 are part of C3 while 7374-7454 are part of the V4 region of env. Areas of prominent RT pausing are indicated by “P” and correspond to those shown in Fig. 6-2 for the A115 donor, D89 acceptor experiment. (B) Table showing the number of recovered transfer products that showed a cross-over within the regions denoted in A for the A115 to D89 experiment. Transfer products were recovered from denaturing polyacrylamide gels (see Fig. 6-2) and amplified by PCR. Products were recovered on native acrylamide gels and inserted into pBSM13+ plasmid that was used to transform *E. coli*. Plasmid DNA recovered from individual colonies was sequenced to determine the cross-over point. A total of 19 and 16 clones from reactions – and + NC, respectively, were sequenced.

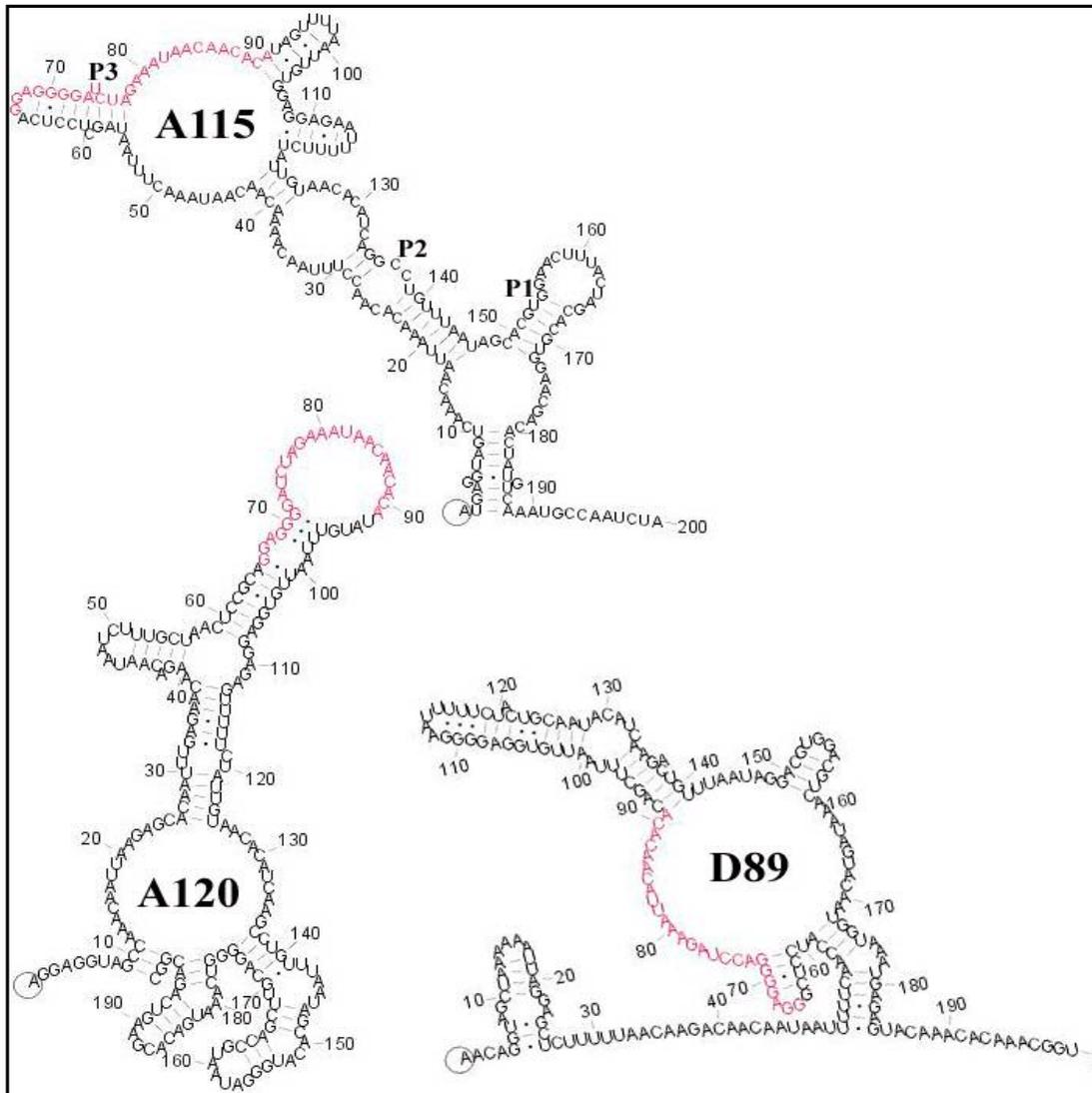


Figure 6-7: Predicted structure of RNAs from clones A115, A120, and D89 over bases 7255-7454 (relative to HIV HXB2 provirus). Fold parameters were from the RNAstructure program and drawing from XRNA. Only the structure with the lowest predicted ΔG is shown. Using default program settings, for A115, A120, and D89, the ΔG values were -34.0, -36.5, -32.8 kcal/mol, resp. The circled base marks the 5' end and corresponds to nt 7255 of the provirus. Red nts denote regions which showed the highest amount of cross-overs in the A115 donor, D89 acceptor experiment performed in vitro (Figure. 6-6B, regions VIII-X). Regions corresponding to the most prominent pause sites (see Figure. 6-2) are denoted P1, P2, and P3. The 56 nt hotspot region identified in cell culture assays corresponds to bases 66-121.

6.5 Discussion

Increased recombination frequencies were observed when A115 RNA was used as donor compared to any other donor RNA in the single cycle assays. This appeared to be related to a unique C3 *env* 'hotspot' for recombination specific to A115. In order to dissect the mechanism(s) of increased strand transfer from the A115 C3 region, a reconstituted *in vitro* reverse transcription system was employed. More RT pausing was observed with the A115 template as compared to A120 template (Figure 6-2, 6-3) and transfer efficiency was 1.5 to 2-fold higher with the A115 donor template compared to the A120 template. This did not result from overall greater homology between A115 and D89 vs. A120 and D89 as in both cases homology was essentially the same at between 66-69%. Preliminary data suggests that specific RNA secondary structures in the C3 region of A115 RNA may be responsible for the preferential strand transfer in this region and hence the observed increase in recombination frequency. Shown in Figure 6-7 are the folded structures for D89, A115 and A120 in the C3 'hotspot' region. The red colored bases denote regions X-XIII in Figure. 6-6, B where transfer was greatest. All three RNAs form a similar structure in this region with a single stranded region that is followed by a small hairpin. Of note is the pause site (P3, see Figure. 6-2) in the stem region of A115. No pause was detected in A120 in this region. Our hypothesis is that RT pausing at P3 gives the D89 acceptor template time to bind to the single stranded region that precedes the pause site. The acceptor can then displace the donor leading to recombination. The single stranded nature of this region probably makes it easier for the acceptor to rapidly bind (see Chapter 5). There is also relatively high homology between A115, A120 and D89 in the red base zone (92% identity in a 26 nucleotide stretch) and

this would make it easier for the acceptor and DNA to hybridize. The lack of a pause on A120 may be responsible for the lower level of recombination. Experiments to test this mechanism by making mutations that alter the level of homology in the red base region and mutations that affect the level of pausing at P3 are planned.

Chapter 7 General Discussion

The process of reverse transcription is one of the well-understood steps in the life cycle of retroviruses (218). Much of what is known about the process today was learned by identifying replication intermediates synthesized in endogenous reactions or in reconstituted *in vitro* reactions and also by analysis of viral DNA isolated from infected cells (61-64, 195). However, there are several details of the process that are not clear. In actively growing cells, under normal circumstances, reverse transcription proceeds in an uninterrupted fashion through the steps discussed in Chapter I and DNA synthesis is essentially completed in the cytoplasm. The process of reverse transcription is generally completed within 8 to 12 hours (218). HIV virion cores are disrupted shortly after virus-cell fusion and reverse transcription proceeds to completion in a large nucleoprotein complex. Disassembly of the core is essential to allow reverse transcription to progress as the Lv-1 restriction phenotype (targets p2 domain of p24), which prevents uncoating of the viral core inhibits reverse transcription (291, 292). The ribonucleoprotein complex within which reverse transcription takes place provides a particularly favorable environment for reverse transcription and appears to resemble, at least superficially the virion core (293). Biochemical, confocal and EM studies show that this complex contains the diploid RNA genome and proteins such as RT, NC, IN, MA (294, 295) and probably Vpr (65). NC enhances viral DNA synthesis by its chaperone activity. The role of other proteins of the complex in reverse transcription remains unclear. Other proteins of HIV-1 that are not known to be part of the ribonucleoprotein complex have been implicated at least indirectly, to have a role in reverse transcription. It has been reported that Nef-

deficient viruses produce reduced amounts of viral DNA (296) and that Tat and Vif have a role in reverse transcription (207, 297, 298).

Host factors may have a role in reverse transcription *in vivo* and several host proteins are known to associate with and be packaged in HIV virions (299). One particular host protein, cyclophilin A, which binds to HIV-1 capsid protein (300) is required for proper entry or uncoating (301, 302). Cyclosporin drugs, which bind to cyclophilins and disrupt the cyclophilin interaction with CA block some early events before viral DNA synthesis begins (303). Other host factors may also play a role. It is possible that some components of the virion are joined by or interact with, these host factors from either the cell that produced the virus or the infected cell. A thorough characterization of the viral nucleoprotein complex is extremely difficult because of the relatively low abundance of these complexes compared to various host-cell nucleoprotein assemblies and also by the low ratio of particle to infectivity in retroviruses. Infected cells contain a relatively large proportion of products from defective virions as most HIV virions, like those of many other viruses, are defective. It is not known whether replication is aborted because of loss of such essential components from the complex or for other reasons. This is because it is not possible to separate replication-competent viral nucleoprotein complexes from replication-defective ones.

Complete viral DNA can be synthesized in purified virions in “endogenous reactions” (ERT), although inefficiently (195-198). In these reactions, the purified virions are permeabilized with mild detergent and incubated with deoxyribonucleotides. Small

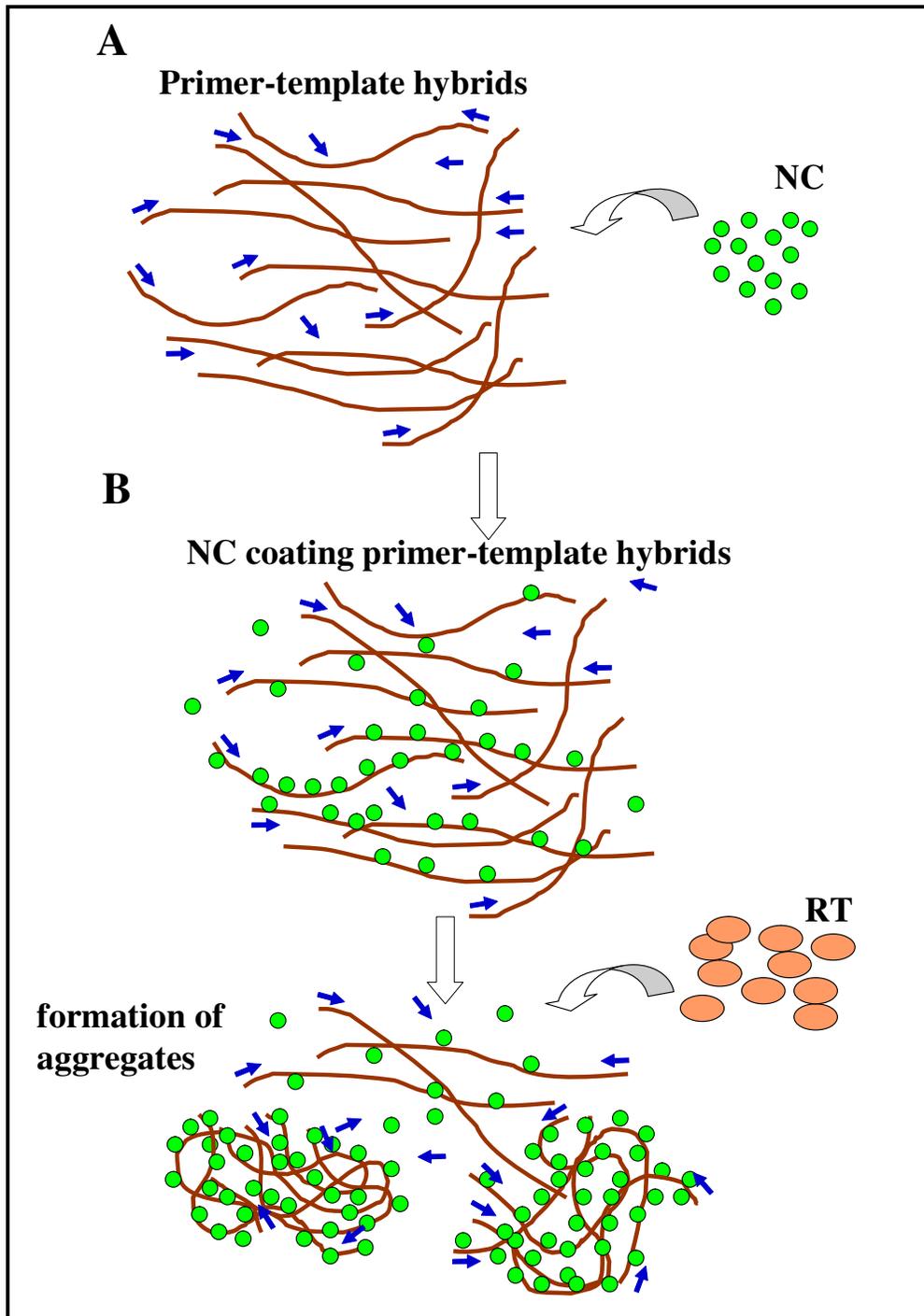
amounts of full-length viral DNA can be synthesized in these reactions, although the reaction is inefficient, it demonstrates that all the components except dNTPs required for reverse transcription reaction are present in the virion (195-198). This also indicates that the absence of dNTPs may be the primary reason why reverse transcription is not initiated in virions. Synthesis of complete reverse transcripts in endogenous reactions is very sensitive to the concentration of detergents present in these reactions (196, 197). With very little detergent, overall reverse transcription is poor whereas with too much detergent, early reverse transcription intermediates are formed but strand transfer is inefficient and plus-strand DNA synthesis does not occur. This sensitivity to detergent suggests that an intact core is required for efficient reverse transcription. However, it is not known whether the core serves to confine the components of reverse transcription at high concentrations or whether the structure of the core is important in promoting reverse transcription. It is also not known whether such detergent treated cores mimic the intracellular cores that carry out reverse transcription. Reverse transcription can also be carried out in extracellular virions without application of detergents. These are referred to as “natural endogenous reactions” (NERT) (199-201). In NERT reactions, the amphipathic domains at the C-terminal end of gp41 make the envelope naturally permeable to dNTPs (201). There are significant differences in reverse transcription efficiency in these endogenous reaction systems and *in vivo*. In ERT and NERT reactions, even under conditions most conducive to reverse transcription, only a few percent of DNA is full-length double-stranded (198) whereas *in vivo* most of the viral DNA seen is full-length. A detailed understanding of how the process of reverse transcription is carried out so efficiently in the host cell and of the various factors and

conditions involved in the process might lead to the development of more effective anti-HIV therapies and also would be an integral part of our understanding of the ways in which genetic information is passed on to future generations in these viruses.

The main area of study in this thesis is to generate conditions *in vitro* that allow efficient synthesis of long reverse transcription products from genomic RNA of human immunodeficiency virus and to analyze the mechanism(s) that leads to the synthesis of these long products. Reverse transcription can be carried out in reconstituted *in vitro* reactions. However, the process is inefficient with most products being about a few hundred to a few thousand bases in length even when RNA templates of several thousand bases are used (184). A previous report has shown that at conditions of saturation of NC binding sites, i.e., one NC molecule for approximately seven nucleotides *in vitro*, full-length DNA was synthesized when an RNA template of 874 bases was used (209). In this study, conditions were optimized that produced DNA products up to 4 kb from genomic RNA of HIV at relatively high efficiency. These products were synthesized in high molecular weight aggregates formed in presence of NC. Internal strand transfer was found to be the molecular mechanism involved in the synthesis of long DNA products. Figure 7-1 represents the proposed model for the synthesis of long reverse transcription products *in vitro*.

There was a decrease in primer extension in the presence of NC at saturating concentrations with almost all products synthesized being full-length. There are several possible explanations for this decrease in synthesis. NC is a chaperone protein with weak

Proposed Model for Synthesis of long DNA products *in vitro*



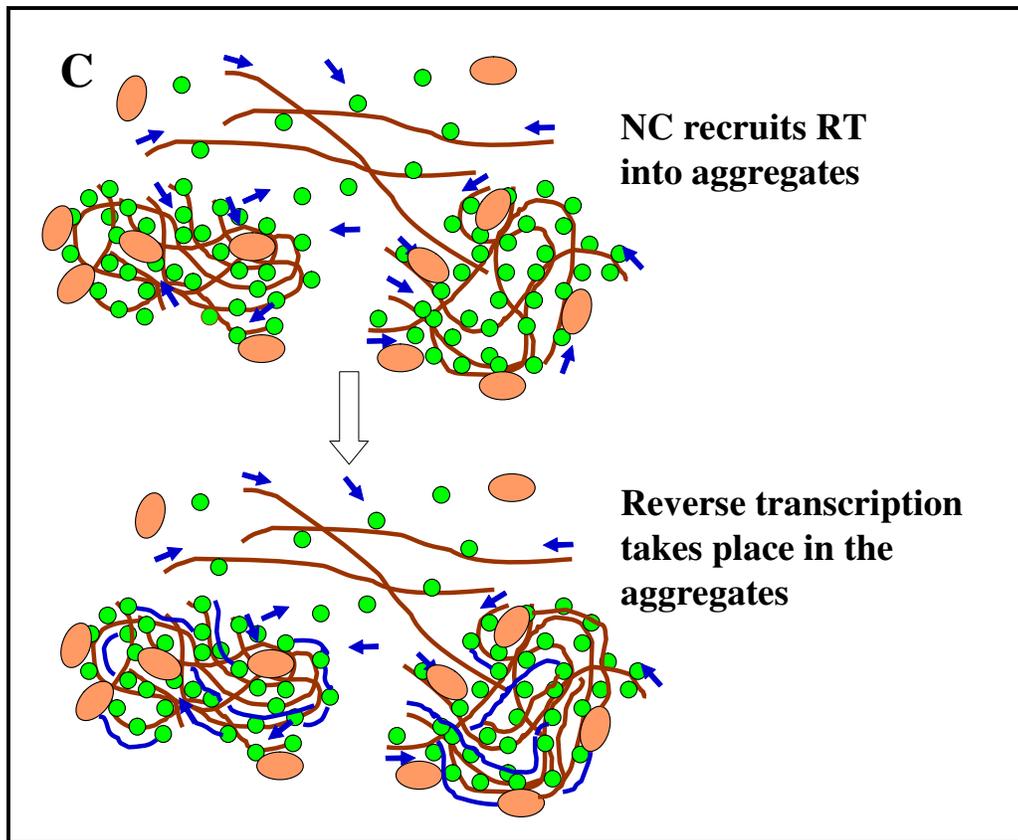


Figure 7-1: Proposed model for synthesis of long DNA products *in vitro*

A represents the first step in the synthesis of long DNA products *in vitro* which is making primer-template hybrids. Not all primers and templates form hybrids. Some of the primers and templates remain unhybridized in these reactions. Nucleocapsid protein (NC) is then added to the reaction.

B When NC is added to the reactions, NC coats the primer-template hybrids. All the nucleic acids in the reaction are not completely coated with NC although the amount of NC added to these reactions is sufficient to coat all nucleic acids. This is due to the limitations in protein and nucleic acid solubility *in vitro*. Those primer-template hybrids that are completely coated with NC enter into high molecular weight aggregates. Formation of these aggregates creates the concentrated and condensed environment conducive to the synthesis of long reverse transcription products. Reverse transcriptase enzyme (RT) is then added to these reactions.

C When RT is added to these reactions, NC recruits RT into these high molecular weight aggregates. Reverse transcription then takes place in these aggregates and involves the process of internal strand transfer which leads to the formation of long DNA products in the aggregates. The high molecular weight aggregates formed promote synthesis of long reverse transcription products by concentrating the nucleic acids, NC and RT into a smaller space thereby mimicking the role of the capsid environment within the host cell which is crucial for efficient reverse transcription in the cell.

helix destabilizing activity and therefore NC may destabilize primer template hybrids thereby having an effect on total DNA synthesis. Consistent with this hypothesis, in experiments using a longer primer (50 nucleotides in length) instead of a shorter primer (20 nucleotides in length) an improvement in the level of synthesis was observed (data not shown). Also, in experiments with template in large excess over primer showed greater levels of extension (Chapter 2). This suggests that inhibition in the presence of NC can in part be overcome by increasing primer binding strength or the number of primer binding sites. Although NC is present at saturating concentrations enough for 1 NC molecule per 7 nucleotides, in practice, not all nucleic acids present in the reaction pool may be completely coated with NC and complete coating, as is the case in virion RNA, may be required for synthesis of full-length DNA. In the virion, NC and RT are condensed within the virus core and present at mM concentrations (304). Limitations in protein and nucleic acid solubility make it essentially impossible to reproduce these concentrations in the test tube. A method to ensure complete coating of all RNA in the reaction would eliminate this limitation, if it were one, in the synthesis of long DNA products *in vitro*. Analysis of the mechanism for synthesis of long reverse transcription products reveal the formation of high molecular weight aggregates in the presence of NC, early in reverse transcription, which play an important role in long DNA synthesis. These aggregates may promote long DNA synthesis by concentrating the components of reverse transcription thereby perhaps mimicking the role of reverse transcription complexes in the host cell. The fact that the aggregate can efficiently produce long DNA products suggests that the condensing effect of the virion environment may be a key for efficient reverse transcription and cellular factors may not be required. A decrease in synthesis in

the presence of NC could also be due to the inability or failure of all the RNA templates or primers to get trapped in these aggregates. If this were the case, then this could explain the decrease in synthesis with NC especially since these aggregates seem to mimic virion core environments that are absolutely required for efficient reverse transcription *in vivo* (191). Experiments are currently being conducted to determine what proportion of the template, primer, and RT is captured in the aggregate complex during normal *in vitro* reactions. Preliminary results suggest that most of the RT and much of the primer are not in the aggregate while nearly all the template is. This separation of primer and template could explain the low level of primer extension with NC.

Repeated attempts to synthesize reverse transcription products longer than 4 kb failed and primer extension was severely inhibited in reactions with the 4 kb template (Chapter 2). Limitations to synthesis of very long DNAs *in vitro* could be due to inability to synthesize sufficient long RNA templates by *in vitro* run-off transcription methods (219) or that the synthesis conditions used *in vitro* are not conducive to synthesis of genome length DNAs. It is possible that some viral or cellular component that is essential for synthesis of genome length DNAs was not included in these *in vitro* reactions or that a combination of these factors was responsible[39].

An *in vitro* system generated here for the synthesis of 1.6 kb could be designed to mimic all the steps of retroviral reverse transcription by using an RNA template of 1.6 kb that contains all sequence elements important for reverse transcription. These include the direct repeat regions, termed R that lie at the 5' and 3' ends of the RNA immediately

adjacent to the 5' cap and 3' poly (A) tail, sequences unique to each end of the viral RNA termed U5 and U3, the primer binding site (PBS) and the polypurine tract (PPT) that are required either to position or to become the RNA primers for minus- and plus-strand DNA synthesis. A tRNA^{Lys3} primer can be used to initiate reverse transcription. A system like this would more closely mimic retroviral reverse transcription in the cell.

An in vitro system that mimics intracellular retroviral replication would enhance our understanding of the replication process and could also potentially be used to screen reverse transcription inhibitors particularly inhibitors of NC, RT and recombination. It may also be possible to use these types of systems to improve the synthesis of cDNA from RNA. This would depend on the source of the RNA. For example it probably would not be particularly useful for producing cDNAs from pooled cellular mRNA. Because such a pool would have several different mRNAs with unique sequences, a mechanism dependent on homologous strand transfer would be unlikely to improve synthesis, especially for those mRNA not present at high concentrations. However, for those cases where a single RNA species is being used (for example the RNA genome from a particular virus), a strand transfer mechanism could allow efficient synthesis of longer RNA segments. This could abrogate the need for cloning long genomes by combining smaller segments.

A major obstacle facing the development of successful vaccines or drugs against HIV is genetic recombination and high mutation rates of the virus. Currently, the HIV strains are classified into three groups; M, N and O of which group M causes over 99% of

the world's AIDS cases. Genetic recombination and mutations have resulted in the emergence of diverse HIV subtypes or classes (305) among Group M which is currently further classified into 10 subtypes or clades: A-K. In patients with chronic infections, the virus causing the initial infection can undergo *in vivo* evolution to produce genetically distinct variants resulting in a quasispecies population (306, 307). This has important consequences in the pathogenesis and treatment of HIV-1 infection. Reports show that recombination takes place between A and D subtypes (308) and also between A and C subtypes (309) in individuals co-infected with these subtypes. It has been shown that eight of the ten subtypes of group M have recombined to produce genetic variants. It is difficult to study intersubtype recombination in *in vivo* experiments or animal models. *In vitro* and *ex vivo* (infections of tissue culture cells) methods are alternatives that can help identify 'hotspots' for recombination. For example, gp120, the most likely vaccine target among HIV proteins, is made of 5 variable (V) regions and 4 constant (C) regions, V1-V5 and C1-C4. 'Hotspots' for recombination have been identified in the C2 region by *in vitro* (261) and *ex vivo* (260) methods. Analysis of HIV isolates shows that subgroups A combine with C and D in the C2 region of gp120 (283) consistent with *in vitro* and *ex vivo* data. This results in genetic reallocation of the variable regions, which helps the virus escape neutralizing host antibodies (310). This also poses problems for eliciting broad range immune responses necessary for a successful vaccine. The second area of study in this thesis involves characterization of a recombination hotspot in the C3 region between A to D subtype recombinants isolated from Ugandan patients by analyzing strand transfer events using *in vitro* reconstituted HIV-1 reverse transcription assays. The exact mechanism for increased strand transfer in a particular A subtype, A115 specific C3

'hotspot' region needs to be determined. Analysis of the mechanism of recombination between intersubtypes of HIV-1 at the recombination 'hotspots' may provide useful information while designing HIV vaccine constructs.

BIBLIOGRAPHY

1. Coffin, J.M. *Structure and classification of retroviruses*. In *The retroviridae* (ed. J. A. Levy), 1992. p.19-49. Plenum Press, New York.
2. Murphy, F.A., C.M. Fauquet, D.H.L. Bishop, S.A. Ghabrial, A.W. Jarvis, G.P. Martelli, M.A. Mayo and M.D. Summers, eds. *Virus taxonomy: The classification and nomenclature of viruses*. Retroviridae, 1994. Springer-Verlag, Vienna.
3. No authors listed. *From the Centers for Disease Control and Prevention. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults*. JAMA, 1993. **269**(6): p. 729-730.
4. Wain-Hobson S., Vartanian J.P., Henry M., Chenciner N., Cheynier R., Delassus S., Martins L.P., Sala M., Nugeyre M.T., Guetard D., *et al.* *LAV revisited: origins of the early HIV-1 isolates from Institut Pasteur*. Science, 1991. **252**(5008): p. 961-965.
5. Barre-Sinoussi F., Chermann J.C., Rey F., Nugeyre M.T., Chamaret S., Gruest J., Dauguet C., Axler-Blin C., Vezinet Brun F., Rouzioux C., Rozenbaum W., Montagnier L. *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)*. Science 1983. **220**(4599): p. 868-871.
6. Gallo R.C., Salahuddin S.Z., Popovic M., Shearer G.M., Kaplan M., Haynes B.F., Pkaker T.J., Redfield R., Oleske J., Safai B., *et al.* *Frequent detection and*

- isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science, 1984. 224(4648): p. 500-503.*
7. Levy J.A., Hoffman A.D., Kramer S.M., Landis J.A., Shimabukuro J.M., Oshiro L.S. *Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. Science, 1984. 225(4664): p. 840-842.*
 8. Clavel F., Guetard D., Brun-Vezinet F., Chamaret S., Rey M.A., Santos-Ferreira M.O., Laurent A.G., Dauguet C., Katlama C., Rouzioux C., *et al. Isolation of a new human retrovirus from West African patients with AIDS. Science, 1986. 233(4761): p. 343-346.*
 9. Pepin J., Morgan G., Dunn D., Gevao S., Mendy M., Gaye I., Scollen N., Tedder R., Whittle H. *HIV-2-induced immunosuppression among asymptomatic West African prostitutes: evidence that HIV-2 is pathogenic, but less so than HIV-1. AIDS, 1991. 5(10): p. 1165-1172.*
 10. Markovitz D.M. *Infection with the human immunodeficiency virus type 2. Ann. Intern. Med., 1993. 118(3): p. 211-218.*
 11. Myers, G., B. Korber, S. Wain-Hobson, K.-T. Jeang, L.E. Henderson and G.N. Pavlakis, Eds. 1994. *Human retroviruses and AIDS 1994.* Los Alamos National Laboratory, Los Alamos, New Mexico.
 12. Louwagie J., Janssens W., Mascola J., Heyndrickx L., Hegerich P., vander Groen G., McCutchan F.E., Burke D.S. *Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. J. Virol., 1995, 69(1): p. 263-271.*

13. Fauci, A.S. and H.C. Lane. *The acquired immunodeficiency syndrome (AIDS)*. In Harrison's principles of internal medicine (ed. J. D. Wilson *et al*), 1991. p. 1402 - 1410. McGraw-Hill, New York.
14. de Vincenzi I. *A longitudinal study of human immunodeficiency virus transmission by heterosexual partners. European Study Group on Heterosexual Transmission of HIV*. N. Engl. J. Med., 1994. **331**(6): 341-346.
15. Holmberg, S.D. *Risk factors for sexual transmission of human immunodeficiency virus*. In AIDS: Etiology, diagnosis, treatment and prevention (ed. V. T. deVita *et al.*), 1997. p. 569-571. Lippincott-Raven, Philadelphia.
16. Scarlatti G. *Paediatric HIV infection*. Lancet, 1996. **348**(9031): p. 863-868.
17. Gao F., Bailes E., Robertson D.L., Chen Y., Rodenburg C.M., Michael S.F., Cummins L.B., Arthur L.O., Peeters M., Shaw G.M., Sharp P.M., Hahn B.H. *Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes*. Nature, 1999. **397**(6718): p. 436-441.
18. Hahn B.H., Shaw G.M., De Cock K.M., Sharp P.M. *AIDS as a zoonosis: scientific and public health implications*. Science, 2000. **287**(5453): p. 607-614.
19. Korber B., Muldoon M., Theiler J., Gao F., Gupta R., Lapedes A., Hahn B.H., Wolinsky S., Bhattacharya T. *Timing the ancestor of the HIV-1 pandemic strains*. Science, 2000. **288**(5472): p. 1789-1796.
20. Gao F., Yue L., White A.T., Pappas P.G., Barchue J., Hanson A.P., Greene B.M., Sharp P.M., Shaw G.M., Hahn B.H. *Human infection by genetically diverse SIVSM-related HIV-2 in West Africa*. Nature., 1992. **358**(6386): p. 495-499.

21. Hirsch V.M., Olmsted R.A., Murphey-Corb M., Purcell R.H., Johnson P.R. *An African primate lentivirus (SIVsm) closely related to HIV-2*. Nature, 1989. **339**(6223): p. 389-392.
22. UNAIDS, *Global summary of the HIV/AIDS epidemic*. 2005.
23. Dalgleish A.G, Beverley P.C., Clapham P.R., Crawford D.H., Greaves M.F., Weiss R.A. *The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus*. Nature, 1984. **312**(5996): p. 763-767.
24. Klatzmann D., Champagne E., Chamaret S., Gruet J., Guetard D., Hercend T., Gluckman J.C., Montagnier L. *T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV*. Nature, 1984. **312**(5996): p. 767-768.
25. Doyle C., Strominger J.L. *Interaction between CD4 and class II MHC molecules mediates cell adhesion*. Nature, 1987. **330**(6145): p. 256-259.
26. Gay D., Maddon P., Sekaly R., Talle M.A., Godfrey M., Long E., Goldstein G., Chess L., Axel R., Kappler J., et al. *Functional interaction between human T-cell protein CD4 and the major histocompatibility complex HLA-DR antigen*. Nature, 1987. **328**(6131): p. 626-629.
27. Alkhatib G., Combadiere C., Broder C.C., Feng Y., Kennedy P.E., Murphy P.M., Berger E.A. *CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1*. Science 1996. **272**(5270): p. 1955-1958.
28. Deng H., Liu R., Ellmeier W., Choe S., Unutmaz D., Burkhart M., Di Marzio P., Marmon S., Sutton R.E., Hill C.M., Davis C.B., Peiper S.C., Schall T.J., Littman D.R., Landau N.R. *Identification of a major co-receptor for primary isolates of HIV-1*. Nature, 1996. **381**(6584): p. 661-666.

29. Dragic T., Litwin V., Allaway G.P., Martin S.R., Huang Y., Nagashima K.A., Cayanan C., Maddon P.J., Koup R.A., Moore J.P., Paxton W.A. *HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5*. *Nature*, 1996. **381**(6584): p. 667-673.
30. Rosenberg Z.F. and A.S. Fauci. *Immunopathogenesis of HIV infection*. In *AIDS: Etiology, diagnosis, treatment and prevention*, 3rd edition (ed. V. T. DeVita, Jr. *et al.*), 1992. p. 61-76. J.B.Lippincott, Philadelphia.
31. Rosenberg Z.F., Fauci A.S. *The immunopathogenesis of HIV infection*. *Adv. Immunol.*, 1989. **47**: p. 377-431.
32. Narayan O., Kennedy-Stoskopf S., Zink M.C. *Lentivirus-host interactions: lessons from visna and caprine arthritis-encephalitis viruses*. *Ann. Neurol.*, 1988. **23** Suppl.: S95-100.
33. Ameisen J.C., Capron A. *Cell dysfunction and depletion in AIDS: the programmed cell death hypothesis*. *Immunol. Today*, 1991. **12**(4): p. 102-105.
34. Groux H., Torpier G., Monte D., Mouton Y., Capron A., Ameisen J.C. *Activation-induced death by apoptosis in CD4+ T cells from human immunodeficiency virus-infected asymptomatic individuals*. *J. Exp. Med.*, 1992. **175**(2): p. 331-340.
35. Pantaleo G., Graziosi C., Demarest J.F., Butini L., Montroni M., Fox C.H., Orenstein J.M., Kotler D.P., Fauci A.S. *HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease*. *Nature*, 1993. **362**(6418): p. 355-358.

36. Ziegler J.L., Stites D.P. *Hypothesis: AIDS is an autoimmune disease directed at the immune system and triggered by a lymphotropic retrovirus.* Clin. Immunol. Immunopathol., 1986. **41**(3): p. 305-313.
37. Fauci A.S. *The human immunodeficiency virus: infectivity and mechanisms of pathogenesis.* Science, 1988. **239**(4840): p. 617-622.
38. Lang W., Perkins H., Anderson R.E., Royce R., Jewell N., Winkelstein W. Jr. *Patterns of T lymphocyte changes with human immunodeficiency virus infection: from seroconversion to the development of AIDS.* J. Acquir. Immune Defic. Syndr., 1989. **2**(1): p. 63-69.
39. Crowe S.M., Carlin J.B., Stewart K.I., Lucas C.R., Hoy J.F. *Predictive value of CD4 lymphocyte numbers for the development of opportunistic infections and malignancies in HIV-infected persons.* J. Acquir. Immune Defic. Syndr., 1991. **4**(8): p. 770-776.
40. Rutherford G.W., Lifson A.R., Hessol N.A., Darrow W.W., O'Malley P.M., Buchbinder S.P., Barnhart J.L., Bodecker T.W., Cannon L., Doll L.S., *et al.* *Course of HIV-1 infection in a cohort of homosexual and bisexual men: an 11 year follow up study.* Br. Med. J., 1990. **301**(6762): p. 1183-1188.
41. Moss A.R., Bacchetti P. *Natural history of HIV infection.* AIDS, 1989. **3**(2): p. 55-61.
42. Lemp G.F., Payne S.F., Rutherford G.W., Hessol N.A., Winkelstein W. Jr, Wiley J.A., Moss A.R., Chaisson R.E., Chen R.T., Feigal D.W. Jr., *et al.* *Projections of AIDS morbidity and mortality in San Francisco.* JAMA, 1990. **263**(11): p. 1497-1501.

43. Lemp G.F., Payne S.F., Neal D., Temelso T., Rutherford G.W. *Survival trends for patients with AIDS*. JAMA, 1990. **263**(3): p. 402-406.
44. Seage G.R. 3rd, Oddleifson S., Carr E., Shea B., Makarewicz-Robert L., van Beuzekom M., De Maria A. *Survival with AIDS in Massachusetts, 1979 to 1989*. Am. J. Public Health, 1993. **83**(1): p. 72-78.
45. Clark S.J., Saag M.S., Decker W.D., Campbell-Hill S., Roberson J.L., Veldkamp P.J., Kappes J.C., Hahn B.H., Shaw G.M. *High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection*. N. Engl. J. Med., 1991. **324**(14): p. 954-960.
46. Daar E.S., Moudgil T., Meyer R.D., Ho D.D. *Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection*. N. Engl. J. Med., 1991. **324**(14): p. 961-964.
47. Cooper D.A., Imrie A.A., Penny R. *Antibody response to human immunodeficiency virus after primary infection*. J. Infect. Dis., 1987. **155**(6): p. 1113-1118.
48. Gaines H., von Sydow M., Sonnerborg A., Abert J., Czajkowski J., Pehrson P.O., Chiodi F., Moberg L., Fenyo E.M., Asjo B., *et al.* *Antibody response in primary human immunodeficiency virus infection*. Lancet, 1987. **1**(8544): p. 1249-1253.
49. Duffalo M.L., James C.W. *Enfuvirtide: a novel agent for the treatment of HIV-1 infection*. Ann. Pharmacother., 2003. **37**(10): p.1448-1456.
50. UNAIDS, *HIV vaccine trial results are an important step forward in developing an effective vaccine, say WHO and UNAIDS*. Press release, 2003.

51. VaxGen, *VaxGen announces results of its phase III HIV vaccine trial in Thailand: vaccine fails to meet endpoint*. Press release, 2003.
52. Johnston P.M. and Flores J., Division of AIDS, NIAID. *Progress in HIV vaccine development*. Reprinted from *Current Opinion in Pharmacology*, 2001. **1**(5). p. 504-510.
53. Coffin, J.M., S.H. Hughes, and H.E. Varmus, *Retroviral virions and genomes*. *Retroviruses*, 1997. p. 27-70. Cold Spring Harbor Laboratory Press.
54. Hu W.S., Temin H.M. *Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination*. *Proc. Natl. Acad. Sci. U. S. A.*, 1990. **7**(4): p.1556-1560.
55. Sattentau Q.J., Moore J.P, Vignaux F., Traincard F., Poignard P. *Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding*. *J. Virol.*, 1993. **67**(12): p. 7383-7393.
56. Sattentau Q.J., Moore J.P. *The role of CD4 in HIV binding and entry*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 1993. **342**(1299): p. 59-66.
57. Kélian M ., Jungerwirth S. *Mechanisms of enveloped virus entry into cells*. *Mol. Biol. Med.*, 1990. **7**(1): p. 17-31.
58. Pauza C.D. *The endocytic pathway for human immunodeficiency virus infection*. *Adv. Exp. Med. Biol.*, 1991. **300**. p.111-138; discussion 139-44.
59. Stein B.S., Gowda S.D., Lifson J.D., Penhallow R.C., Bensch K.G., Engleman E.G. *pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane*. *Cell.*, 1987. **49**(5): p. 659-668.

60. McClure M.O., Marsh M., Weiss R.A. *Human immunodeficiency virus infection of CD4-bearing cells occurs by a pH-independent mechanism.* EMBO J., 1988. 7(2): p. 513-518.
61. Weinberg R.A. *Structure of the intermediates leading to the integrated provirus.* Biochim. Biophys. Acta., 1977, 473(1): p. 39-55.
62. Varmus H.E., Heasley S., Kung H.J., Oppermann H., Smith V.C., Bishop J.M., Shank P.R. *Kinetics of synthesis, structure and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells.* J. Mol. Biol., 1978,120(1): p.55-82.
63. Boone L.R., Skalka A.M. *Viral DNA synthesized in vitro by avian retrovirus particles permeabilized with melittin. I. Kinetics of synthesis and size of minus- and plus-strand transcripts.* J. Virol., 1981. 37(1): p.109-116.
64. Borroto-Esoda K., Boone L.R. *Equine infectious anemia virus and human immunodeficiency virus DNA synthesis in vitro: characterization of the endogenous reverse transcriptase reaction.* J. Virol., 1991. 65(4): p.1952-1959.
65. Fassati A., Goff S.P. *Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1.* J. Virol, 2001. 75(8): p. 3626-3635.
66. Priel E., Aflalo E., Seri I., Henderson L.E., Arthur L.O., Aboud M., Segal S., Blair D.G. *DNA binding properties of the zinc-bound and zinc-free HIV nucleocapsid protein: supercoiled DNA unwinding and DNA-protein cleavable complex formation.* FEBS Lett., 1995. 362(1): p. 59-64.

67. Lapadat-Tapolsky M., Pernelle C., Borie C., Darlix J.L. *Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1*. Nucleic Acids Res., 1995. **23**(13): p. 2434-2441.
68. Temin H.M., Kang C.Y., Mizutani S. *Endogenous RNA-directed DNA polymerase activity in normal cells*. Johns Hopkins Med. J. Suppl., 1973.**2**: p. 141-156.
69. Coffin J.M., Haseltine W.A. *Terminal redundancy and the origin of replication of Rous sarcoma virus RNA*. Proc. Natl. Acad. Sci. U. S. A., 1977. **74**(5):1908-1912.
70. Cen S., Huang Y., Khorchid A., Darlix J.L., Wainberg M.A., Kleiman L. *The role of Pr55(gag) in the annealing of tRNA³Lys to human immunodeficiency virus type 1 genomic RNA*. J. Virol., 1999. **73**(5): p. 4485-4488.
71. Yu H., Jetzt A.E., Ron Y., Preston B.D., Dougherty J.P. *The nature of human immunodeficiency virus type 1 strand transfers*. J. Biol. Chem., 1998. **273**(43): p. 28384-28391.
72. Guo J., Wu T., Anderson J., Kane B.F., Johnson D.G., Gorelick R.J., Henderson L.E., Levin J.G. *Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer*. J. Virol., 2000. **74**(19): p. 8980-8988.
73. Charneau P., Alizon M., Clavel F. *A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication*. J. Virol., 1992. **66**(5): p. 2814-2820.

74. Renda M.J., Rosenblatt J.D., Klimatcheva E., Demeter L.M., Bambara R.A., Planelles V. *Mutation of the methylated tRNA (Lys)(3) residue A58 disrupts reverse transcription and inhibits replication of human immunodeficiency virus type 1*. J. Virol., 2001. **75**(20): p. 9671-9678.
75. Hughes S.H., Shank P.R., Spector D.H., Kung H.J., Bishop J.M., Varmus H.E., Vogt P.K., Breitman M.L. *Provirus of avian sarcoma virus are terminally redundant, co-extensive with unintegrated linear DNA and integrated at many sites*. Cell, 1978.**15**(4): p.1397-1410.
76. Brown P.O., Bowerman B., Varmus H.E., Bishop J.M. *Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein*. Proc. Natl. Acad. Sci. U. S. A., 1989. **86**(8): p. 2525-2529.
77. Schroder A.R., Shinn P., Chen H., Berry C., Ecker J.R., Bushman F. *HIV-1 integration in the human genome favors active genes and local hotspots*. Cell, 2002. **110**(4): p. 521-529.
78. Roth M.J., Schwartzberg P.L., Goff S.P. *Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence*. Cell, 1989. **58**(1): p. 47-54.
79. Wain-Hobson S. *Human immunodeficiency virus type 1 quasispecies in vivo and ex vivo*. Curr. Top. Microbiol. Immunol., 1992. **176**: p. 181-193.
80. Clavel F., Hoggan M.D., Willey R.L., Strebel K., Martin M.A., Repaske R. *Genetic recombination of human immunodeficiency virus*. J. Virol., 1989. **63**(3): p. 1455-1459.

81. Jetzt A.E., Yu H., Klarmann G.J., Ron Y., Preston B.D., Dougherty J.P. *High rate of recombination throughout the human immunodeficiency virus type 1 genome*. J. Virol., 2000. **74**(3): p.1234-1240.
82. Zhuang J., Jetzt A.E., Sun G., Yu H., Klarmann G., Ron Y., Preston B.D., Dougherty J.P. *Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots*. J. Virol., 2002. **76**(22): p.11273-11282.
83. Coffin J.M. *Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses*. J. Gen. Virol., 1979. **42**(1): p. 1-26.
84. Hu W.S., Temin H.M. *Retroviral recombination and reverse transcription*. Science, 1990. **250**(4985): p. 1227-1233.
85. Katz R.A., Skalka A.M. *Generation of diversity in retroviruses*. Annu. Rev. Genet., 1990. **24**: p. 409-445.
86. Temin H.M. *Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation*. Proc. Natl. Acad. Sci. U. S. A., 1993. **90**(15): p. 6900-6903.
87. Zhang J., Temin H.M. *Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication*. Science, 1993. **259**(5092): p. 234-238.
88. Levy D.N., Aldrovandi G.M., Kutsch O., Shaw G.M. *Dynamics of HIV-1 recombination in its natural target cells*. Proc. Natl. Acad. Sci. U. S. A., 2004. **101**(12): p. 4204-4209.
89. Onafuwa A., An W., Robson N.D., Telesnitsky A. *Human immunodeficiency virus type 1 genetic recombination is more frequent than that of Moloney murine*

- leukemia virus despite similar template switching rates. J. Virol.*, 2003. **77**(8): p. 4577-4587.
90. Hunter E. *The mechanism for genetic recombination in the avian retroviruses. Curr. Top. Microbiol. Immunol.*, 1978. **79**: p. 295-309.
91. Hu W.S., Temin H.M. *Effect of gamma radiation on retroviral recombination. J. Virol.*, 1992. **66**(7): p. 4457-4463.
92. Xu H., Boeke J.D. *High-frequency deletion between homologous sequences during retrotransposition of Ty elements in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A.*, 1987. **84**(23):8553-8557.
93. Wu W., Blumberg B.M., Fay P.J., Bambara R.A. *Strand transfer mediated by human immunodeficiency virus reverse transcriptase in vitro is promoted by pausing and results in misincorporation. J. Biol. Chem.*, 1995. **270**(1): p. 325-332.
94. Roda R.H., Balakrishnan M., Kim J.K., Roques B.P., Fay P.J., Bambara R.A. *Strand transfer occurs in retroviruses by a pause-initiated two-step mechanism. J. Biol. Chem.*, 2002. **277**(49): p. 46900-46911.
95. Peliska J.A., Benkovic S.J. *Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science*, 1992. **258**(5085): p. 1112-1118.
96. DeStefano J.J., Bambara R.A., Fay P.J. *The mechanism of human immunodeficiency virus reverse transcriptase-catalyzed strand transfer from internal regions of heteropolymeric RNA templates. J. Biol. Chem.*, 1994. **269**(1): p. 161-168.

97. Delviks K.A., Pathak V.K. *Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching.* J. Virol., 1999. **73**(10): p. 7923-7932.
98. Junghans R.P., Boone L.R., Skalka A.M. *Retroviral DNA H structures: displacement-assimilation model of recombination.* Cell, 1982. **30**(1): p. 53-62.
99. Whiting S.H., Champoux J.J. *Strand displacement synthesis capability of Moloney murine leukemia virus reverse transcriptase.* J. Virol., 1994. **68**(8): p. 4747-4758.
100. Baltimore D. *RNA-dependent DNA polymerase in virions of RNA tumor viruses.* Nature, 1970. **226**(5252): p. 1209-1211.
101. Temin H.M., Mizutani S. *RNA-dependent DNA polymerase in virions of Rous sarcoma virus.* Nature, 1970. **226**(5252): p. 1211-1213.
102. Baltimore D., Smoler D.F. *Association of an endoribonuclease with the avian myeloblastosis virus deoxyribonucleic acid polymerase.* J. Biol. Chem., 1972. **247**(22): p. 7282-7287.
103. Shimada M., Hosaka H., Takaku H., Smith J.S., Roth M.J., Inouye S., Inouye M. *Specificity of priming reaction of HIV-1 reverse transcriptase, 2'-OH or 3'-OH.* J. Biol. Chem., 1994. **269**(6): p. 3925-3927.
104. Majumdar C., Abbotts J., Broder S., Wilson S.H. *Studies on the mechanism of human immunodeficiency virus reverse transcriptase. Steady-state kinetics, processivity, and polynucleotide inhibition.* J. Biol. Chem., 1988. **263**(30): p. 15657-15665.

105. Battula N., Loeb L.A. *On the fidelity of DNA replication. Lack of exodeoxyribonuclease activity and error-correcting function in avian myeloblastosis virus DNA polymerase.* J. Biol. Chem., 1976. **251**(4): p. 982-986.
106. Boyer J.C., Bebenek K., Kunkel T.A. *Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates.* Proc. Natl. Acad. Sci. U. S. A., 1992. **89**(15): p. 6919-6923.
107. Yu H., Goodman M.F. *Comparison of HIV-1 and avian myeloblastosis virus reverse transcriptase fidelity on RNA and DNA templates.* J. Biol. Chem., 1992. **267**(15): p. 10888-10896.
108. Le Grice S.F., Naas T., Wohlgensinger B., Schatz O. *Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase.* EMBO J., 1991. **10**(12): p. 3905-3911.
109. Hostomsky Z., Hostomska Z., Fu T.B., Taylor J. *Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis.* J. Virol., 1992. **66**(5): p. 3179-3182.
110. Tasara T., Amacker M., Hubscher U. *Intramolecular chimeras of the p51 subunit between HIV-1 and FIV reverse transcriptases suggest a stabilizing function for the p66 subunit in the heterodimeric enzyme.* Biochemistry, 1999. **38**(5): p. 1633-1642.
111. Harris D., Lee R., Misra H.S., Pandey P.K., Pandey V.N. *The p51 subunit of human immunodeficiency virus type 1 reverse transcriptase is essential in loading the p66 subunit on the template primer.* Biochemistry, 1998. **37**(17): p. 5903-5908.

112. Dufour E., El Dirani-Diab R., Boulme F., Fournier M., Nevinsky G., Tarrago-Litvak L., Litvak S., Andreola M.L. *p66/p51 and p51/p51 recombinant forms of reverse transcriptase from human immunodeficiency virus type 1--interactions with primer tRNA(Lys3), initiation of cDNA synthesis, and effect of inhibitors.* Eur. J. Biochem., 1998. **251**(1-2): p. 487-495.
113. Larder B.A., Purifoy D.J., Powell K.L., Darby G. *Site specific mutagenesis of AIDS virus reverse transcriptase.* Nature, 1987. **327**(6124): p. 716-717.
114. Boyer P.L., Ferris A.L., Hughes S.H. *Mutational analysis of the fingers domain of human immunodeficiency virus type 1 reverse transcriptase.* J. Virol., 1992. **66**(12): p. 7533-7537.
115. Mizrahi V. *Analysis of the ribonuclease H activity of HIV-1 reverse transcriptase using RNA-DNA hybrid substrates derived from the gag region of HIV-1.* Biochemistry, 1989. **28**(23): p. 9088-9094.
116. Repaske R., Hartley J.W., Kavlick M.F., O'Neill R.R., Austin J.B. *Inhibition of RNase H activity and viral replication by single mutations in the 3' region of Moloney murine leukemia virus reverse transcriptase.* J. Virol., 1989. **63**(3): p. 1460-1464.
117. Henderson L.E., Bowers M.A., Sowder R.C. 2nd, Serabyn S.A., Johnson D.G., Bess J.W. Jr., Arthur L.O., Bryant D.K., Fenselau C. *Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences.* J. Virol., 1992. **66**(4): p.1856-1865.

118. Tritch R.J., Cheng Y.E., Yin F.H., Erickson-Viitanen S. *Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 gag polyprotein*. J. Virol., 1991. **65**(2): p. 922-930.
119. Darlix J.L., Lapadat-Tapolsky M., de Rocquigny H., Roques B.P. *First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses*. J. Mol. Biol., 1995. **254**(4): p. 523-537.
120. Karpel R.L., Henderson L.E., Oroszlan S. *Interactions of retroviral structural proteins with single-stranded nucleic acids*. J. Biol. Chem., 1987. **262**(11): p. 4961-4967.
121. Lapadat-Tapolsky M., De Rocquigny H., Van Gent D., Roques B., Plasterk R., Darlix J.L. *Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle*. Nucleic Acids Res., 1993. **21**(4): p. 831-839.
122. South T.L., Blake P.R., Sowder R.C. 3rd, Arthur L.O., Henderson L.E., Summers M.F. *The nucleocapsid protein isolated from HIV-1 particles binds zinc and forms retroviral-type zinc fingers*. Biochemistry, 1990. **29**(34): p.7786-7789.
123. You J.C., McHenry C.S. *HIV nucleocapsid protein. Expression in Escherichia coli, purification, and characterization*. J. Biol. Chem., 1993. **268**(22): p. 16519-16527.
124. Rein A., Henderson L.E., Levin J.G. *Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication*. Trends Biochem. Sci., 1998. **23**(8): p. 297-301.

125. Bess J.W. Jr, Powell P.J., Issaq H.J., Schumack L.J., Grimes M.K., Henderson L.E., Arthur L.O. *Tightly bound zinc in human immunodeficiency virus type 1, human T-cell leukemia virus type I, and other retroviruses.* J. Virol., 1992. **66**(2): p. 840-847.
126. Mely Y., De Rocquigny H., Morellet N., Roques B.P., Gerad D. *Zinc binding to the HIV-1 nucleocapsid protein: a thermodynamic investigation by fluorescence spectroscopy.* Biochemistr, 1996. **35**(16): p. 5175-5182.
127. Mely Y., Jullian N., Morellet N., De Rocquigny H., Dong C.Z., Piemont E., Roques B.P., Gerard D. *Spatial proximity of the HIV-1 nucleocapsid protein zinc fingers investigated by time-resolved fluorescence and fluorescence resonance energy transfer.* Biochemistry, 1994. **33**(40): p. 12085-12091.
128. Morellet N., Jullian N., De Rocquigny H., Maignet B., Darlix J.L., Roques B.P. *Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR.* EMBO J., 1992. **11**(8): p. 3059-3065.
129. Urbaneja M.A., Wu M., Casas-Finet J.R., Karpel R.L. *HIV-1 nucleocapsid protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity.* J. Mol. Biol., 2002. **318**(3): p. 749-764.
130. Rice W.G., Schaeffer C.A., Harten B., Villinger F., South T.L., Summers M.F., Henderson L.E., Bess J.W. Jr., Arthur L.O., McDougal J.S., *et al.* *Inhibition of HIV-1 infectivity by zinc-ejecting aromatic C-nitroso compounds.* Nature, 1993. **361**(6411): 473-475.

131. Rice W.G., Turpin J.A. *Virus-encoded Zinc Fingers as Targets for Antiviral Chemotherapy*. Rev. Med. Virol., 1996. **6**(4): p. 187-199.
132. Turpin J.A., Terpening S.J., Schaeffer C.A., Yu G., Glover C.J., Felsted R.L., Sausville E.A., Rice W.G. *Inhibitors of human immunodeficiency virus type 1 zinc fingers prevent normal processing of gag precursors and result in the release of noninfectious virus particles*. J. Virol., 1996. **70**(9): p. 6180-6189.
133. South T.L., Blake P.R., Hare D.R., Summers M.F. *C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain*. Biochemistry, 1991. **30**(25): p. 6342-6349.
134. Gorelick R.J., Chabot D.J., Rein A., Henderson L.E., Arthur L.O. *The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent*. J. Virol., 1993. **67**(7): p. 4027-4036.
135. Heath M.J., Derebail S.S., Gorelick R.J., DeStefano J.J. *Differing roles of the N- and C-terminal zinc fingers in human immunodeficiency virus nucleocapsid protein-enhanced nucleic acid annealing*. J. Biol. Chem., 2003. **278**(33): p. 30755-30763.
136. Gorelick R.J., Nigida S.M. Jr., Bess J.W. Jr., Arthur L.O., Henderson L.E., Rein A. *Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA*. J. Virol., 1990. **64**(7): p. 3207-3211.
137. South T.L., Summers M.F. *Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC*. Protein Sci., 1993. **2**(1): p. 3-19.

- 138.Lam W.C., Maki A.H., Casas-Finet J.R., Erickson J.W., Sowder R.C. 2nd, Henderson L.E. *Evidence for stacking interactions between 5-mercurated polyuridylic acid and HIV-1 p7 nucleocapsid protein obtained by phosphorescence and optically detected magnetic resonance (ODMR)*. FEBS Lett., 1993. **328**(1-2): p. 45-48.
- 139.Fisher R.J., Rein A., Fivash M, Urbaneja M.A., Casas-Finet J.R., Medaglia M., Henderson L.E. *Sequence-specific binding of human immunodeficiency virus type I nucleocapsid protein to short oligonucleotides*. J. Virol., 1998. **72**(3):1902-1909.
- 140.Vuilleumier C., Bombarda E., Morellet N., Gerard D., Roques B.P., Mely Y. *Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study*. Biochemistry, 1999. **38**(51): p. 16816-16825.
- 141.Amarasinghe GK, De Guzman RN, Turner RB, Chancellor KJ, Wu ZR, Summers MF. *NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition*. J. Mol. Biol., 2000. **301**(2): p. 491-511.
- 142.De Guzman R.N, Wu Z.R., Stalling C.C., Pappalardo L.,Borer P.N., Summers M.F. *Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element*. Science, 1998. **279**(5349): p. 384-388.
- 143.Chertova E.N., Kane B.P., McGrath C., Johnson D.G., Sowder R.C. 2nd, Arthur L.O., Henderson L.E. *Probing the topography of HIV-1 nucleocapsid protein with the alkylating agent N-ethylmaleimide*. Biochemistry, 1998. **37**(51): p. 17890-17897.

- 144.Herschlag D. *RNA chaperones and the RNA folding problem*. J. Biol. Chem., 1995. **270**(36): p. 20871-20874.
- 145.Prats A.C., Housset V., de Billy G., Cornille F., Prats H., Roques B., Darlix J.L. *Viral RNA annealing activities of the nucleocapsid protein of Moloney murine leukemia virus are zinc independent*. Nucleic Acids Res., 1991. **19**(13): p. 3533-3541.
- 146.De Rocquigny H., Ficheux D., Gabus C., Allain B., Fournie-Zaluski M.C., Darlix J.L., Roques B.P. *Two short basic sequences surrounding the zinc finger of nucleocapsid protein NCp10 of Moloney murine leukemia virus are critical for RNA annealing activity*. Nucleic Acids Res., 1993. **21**(4): p. 823-829.
- 147.Osland A., Kleppe K. *Polyamine induced aggregation of DNA*. Nucleic Acids Res., 1977. **4**(3): p. 685-695.
- 148.Pontius B.W., Berg P. *Rapid assembly and disassembly of complementary DNA strands through an equilibrium intermediate state mediated by A1 hnRNP protein*. J. Biol. Chem., 1992. **267**(20): p. 13815-13818.
- 149.Hargittai M.R., Mangla A.T., Gorelick R.J., Musier-Forsyth K. *HIV-1 nucleocapsid protein zinc finger structures induce tRNA(Lys,3) structural changes but are not critical for primer/template annealing*. J. Mol. Biol., 2001. **312**(5): p. 985-997.
- 150.Khan R., Giedroc D.P. *Recombinant human immunodeficiency virus type 1 nucleocapsid (NCp7) protein unwinds tRNA*. J. Biol. Chem., 1992. **267**(10): p. 6689-6695.

151. Remy E., de Rocquigny H., Petitjean P., Muriaux D., Theilleux V., Paoletti J., Roques B.P. *The annealing of tRNA^{3Lys} to human immunodeficiency virus type 1 primer binding site is critically dependent on the NCp7 zinc fingers structure.* J. Biol. Chem., 1998. **273**(9): p. 4819-4822.
152. De Rocquigny H., Gabus C., Vincent A., Fournie-Zaluski M.C., Roques B., Darlix J.L. *Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers.* Proc. Natl. Acad. Sci. U. S. A., 1992. **89**(14): p. 6472-6476.
153. Barat C., Lullien V., Schatz O., Keith G., Nugeyre M.T., Gruninger-Leitch F., Barre-Sinoussi F., LeGrice S.F., Darlix J.L. *HIV-1 reverse transcriptase specifically interacts with the anticodon domain of its cognate primer tRNA.* EMBO J., 1989. **8**(11): p. 3279-3285.
154. Johnson P.E., Turner R.B., Wu Z.R., Hairston L., Guo J., Levin J.G., Summers M.F. *A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription.* Biochemistry, 2000. **39**(31): p. 9084-9091.
155. You J.C., McHenry C.S. *Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription.* J. Biol. Chem., 1994. **269**(50): p.
156. Allain B., Lapadat-Tapolsky M., Berlioz C., Darlix J.L. *Transactivation of the minus-strand DNA transfer by nucleocapsid protein during reverse transcription of the retroviral genome.* EMBO J., 1994. **13**(4): p. 973-981.

157. Guo J., Henderson L.E., Bess J., Kane B., Levin J.G. *Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA*. J. Virol., 1997. **71**(7): p. 5178-5188.
158. Wu T., Guo J., Bess J., Henderson L.E., Levin J.G. *Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems*. J. Virol., 1999. **73**(6): p. 4794-4805.
159. DeStefano J.J. *Interaction of human immunodeficiency virus nucleocapsid protein with a structure mimicking a replication intermediate. Effects on stability, reverse transcriptase binding, and strand transfer*. J. Biol. Chem., 1996. **271**(27): p. 16350-16356.
160. Rodriguez-Rodriguez L., Tsuchihashi Z., Fuentes G.M., Bambara R.A., Fay P.J. *Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro*. J. Biol. Chem., 1995. **270**(25): p. 15005-15011.
161. DeStefano J.J. *Human immunodeficiency virus nucleocapsid protein stimulates strand transfer from internal regions of heteropolymeric RNA templates*. Arch. Virol., 1995. **140**(10): p. 1775-1789.
162. Raja A., DeStefano J.J. *Kinetic analysis of the effect of HIV nucleocapsid protein (NCp) on internal strand transfer reactions*. Biochemistry, 1999. **38**(16): p. 5178-5184.

163. Tsuchihashi Z., Brown P.O. *DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein*. J. Virol., 1994. **68**(9): p. 5863-5870.
164. Bernacchi S., Stoylov S., Piemont E., Ficheux D., Roques B.P., Darlix J.L., Mely Y. *HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence*. J. Mol. Biol., 2002. **317**(3): p. 385-399.
165. Wu W., Henderson L.E., Copeland T.D., Gorelick R.J., Bosche W.J., Rein A., Levin J.G. *Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse transcriptase pausing at a secondary structure near the murine leukemia virus polypurine tract*. J. Virol., 1996. **70**(10): p. 7132-7142.
166. Ji X., Klarmann G.J., Preston B.D. *Effect of human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein on HIV-1 reverse transcriptase activity in vitro*. Biochemistry, 1996. **35**(1): p. 132-143.
167. Lener D., Tanchou V., Roques B.P., Le Grice S.F., Darlix J.L. *Involvement of HIV-1 nucleocapsid protein in the recruitment of reverse transcriptase into nucleoprotein complexes formed in vitro*. J. Biol. Chem., 1998. **273**(50): p. 33781-33786.
168. Dib Hajj F., Khan R., Giedroc D.P. *Retroviral nucleocapsid proteins possess potent nucleic acid strand renaturation activity*. Protein Sci., 1993. **2**(2): p. 231-243.
169. Le Cam E., Coulaud D., Delain E., Petitjean P., Roques B.P., Gerard D., Stoylova E., Vuilleumier C., Stoylov S.P., Mely Y. *Properties and growth*

- mechanism of the ordered aggregation of a model RNA by the HIV-1 nucleocapsid protein: an electron microscopy investigation.* Biopolymers, 1998. **45**(3): p. 217-229.
170. Stoylov S.P., Vuilleumier C., Stoylova E., De Rocquigny H., Roques B.P., Gerard D., Mely Y. *Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein.* Biopolymers, 1997. **41**(3): p. 301-312.
171. Krishnamoorthy G., Roques B., Darlix J.L., Mely Y. *DNA condensation by the nucleocapsid protein of HIV-1: a mechanism ensuring DNA protection.* Nucleic Acids Res., 2003. **31**(18): p. 5425-5432.
172. Carteau S., Batson S.C., Poljak L., Mouscadet J.F., de Rocquigny H., Darlix J.L., Roques B.P., Kas E., Auclair C. *Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg²⁺-dependent DNA integration in vitro.* J. Virol., 1997. **71**(8): p. 6225-6229.
173. Carteau S., Gorelick R.J., Bushman F.D. *Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein.* J. Virol. 1999. **73**(8): p. 6670-6679.
174. Buckman J.S., Bosche W.J., Gorelick R.J. *Human immunodeficiency virus type 1 nucleocapsid zn(2+) fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA.* J. Virol., 2003. **77**(2): p. 1469-1480.
175. Gorelick R.J., Gagliardi T.D., Bosche W.J., Wiltrout T.A., Coren L.V., Chabot D.J., Lifson J.D., Henderson L.E., Arthur L.O. *Strict conservation of the*

- retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences. Virology, 1999. 256(1): p. 92-104.*
- 176.Poon D.T., Wu J., Aldovini A. *Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. J. Virol., 1996. 70(10): p. 6607-6616.*
- 177.Poon D.T., Li G., Aldovini A. *Nucleocapsid and matrix protein contributions to selective human immunodeficiency virus type 1 genomic RNA packaging. J. Virol., 1998. 72(3): p. 1983-1993.*
- 178.Feng Y.X., Copeland T.D., Henderson L.E., Gorelick R.J., Bosche W.J., Levin J.G., Rein A. *HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro. Proc. Natl. Acad. Sci. U. S. A., 1996. 93(15): p. 7577-7581.*
- 179.Fu W., Gorelick R.J., Rein A. *Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. J. Virol., 1994. 68(8): p. 5013-5018.*
- 180.Berthoux L., Pechoux C., Darlix J.L. *Multiple effects of an anti-human immunodeficiency virus nucleocapsid inhibitor on virus morphology and replication. J. Virol., 1999. 73(12): p. 10000-10009.*
- 181.Rice W.G., Supko J.G., Malspeis L., Buckheit R.W. Jr., Clanton D., Bu M., Graham L., Schaeffer C.A., Turpin J.A., Domagala J., Gogliotti R., Bader J.P., Halliday S.M., Coren L., Sowder R.C. 2nd, Arthur L.O., Henderson L.E.

Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS. Science, 1995. **270**(5239): p. 1194-1197.

182. Nagy K., Young M., Baboonian C., Merson J., Whittle P., Oroszlan S. *Antiviral activity of human immunodeficiency virus type 1 protease inhibitors in a single cycle of infection: evidence for a role of protease in the early phase.* J. Virol., 1994. **68**(2): p. 757-765.
183. Rossio J.L., Esser M.T., Suryanarayana K., Schneider D.K., Bess J.W. Jr, Vasquez G.M., Wiltrout T.A., Chertova E., Grimes M.K., Sattentau Q., Arthur L.O., Henderson L.E., Lifson J.D. *Inactivation of human immunodeficiency virus type 1 infectivity with preservation of conformational and functional integrity of virion surface proteins.* J. Virol., 1998. **72**(10): p. 7992-8001.
184. Klarmann G.J., Schaubert C.A., Preston B.D. *Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during polymerization of HIV-1 sequences in vitro.* J. Biol. Chem., 1993. **268**(13): p. 9793-9802.
185. Lori F., di Marzo Veronese F., de Vico A.L., Lusso P., Reitz M.S. Jr., Gallo R.C. *Viral DNA carried by human immunodeficiency virus type 1 virions.* J. Virol., 1992. **66**(8): p. 5067-5074.
186. Trono D. *Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses.* J. Virol., 1992. **66**(8): p. 4893-4900.
187. Zack J.A., Haislip A.M., Krogstad P., Chen I.S. *Incompletely reverse-transcribed human immunodeficiency virus type 1 genomes in quiescent cells can function as intermediates in the retroviral life cycle.* J. Virol., 1992. **66**(3): p. 1717-1725.

- 188.Zhu J., Cunningham J.M. *Minus-strand DNA is present within murine type C ecotropic retroviruses prior to infection.* J. Virol., 1993. **67**(4): p. 2385-2388.
- 189.Zhang H., Bagasra O., Niikura M., Poiesz B.J., Pomerantz R.J. *Intravirion reverse transcripts in the peripheral blood plasma on human immunodeficiency virus type 1-infected individuals.* J. Virol., 1994. **68**(11): p. 7591-7597.
- 190.Grewe C., Beck A., Gelderblom H.R. *HIV: early virus-cell interactions.* J. Acquir. Immune Defic. Syndr., 1990. **3**(10): p. 965-974.
- 191.Forshey B.M., von Schwedler U., Sundquist W.I., Aiken C. *Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication.* J. Virol., 2002. **76**(11): p. 5667-5677.
- 192.Bukrinskaya A., Brichacek B., Mann A., Stevenson M. *Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton.* J. Exp. Med., 1998. **188**(11): p. 2113-2125.
- 193.Negrioni M., Buc H. *Mechanisms of retroviral recombination.* Annu. Rev. Genet., 2001. **35**: p. 275-302.
- 194.Nermut M.V., Fassati A. *Structural analyses of purified human immunodeficiency virus type 1 intracellular reverse transcription complexes.* J. Virol., 2003. **77**(15): p. 8196-8206.
- 195.Haseltine W.A., Kleid D.G., Panet A., Rothenberg E, Baltimore D. *Ordered transcription of RNA tumor virus genomes.* J. Mol. Biol., 1976. **106**(1): p. 109-131.
- 196.Rothenberg E., Baltimore D. *Synthesis of long, representative DNA copies of the murine RNA tumor virus genome.* J. Virol., 1975. **17**(1): p. 168-174.

197. Rothenberg E., Smotkin D., Baltimore D., Weinberg R.A. *In vitro synthesis of infectious DNA of murine leukaemia virus*. Nature, 1977. **269**(5624): p. 122-126.
198. Gilboa E., Goff S., Shields A., Yoshimura F., Mitra S., Baltimore D. *In vitro synthesis of a 9 kbp terminally redundant DNA carrying the infectivity of Moloney murine leukemia virus*. Cell, 1979. **16**(4): p. 863-874.
199. Zhang H., Zhang Y., Spicer T.P., Abbott L.Z., Abbott M., Poiesz B.J. *Reverse transcription takes place within extracellular HIV-1 virions: potential biological significance*. AIDS Res. Hum. Retroviruses, 1993. **9**(12): p. 1287-1296.
200. Zhang H., Dornadula G., Pomerantz R.J. *Endogenous reverse transcription of human immunodeficiency virus type 1 in physiological microenvironments: an important stage for viral infection of nondividing cells*. J. Virol., 1996. **70**(5): p. 2809-2824.
201. Zhang H., Dornadula G., Alur P., Laughlin M.A., Pomerantz R.J. *Amphipathic domains in the C terminus of the transmembrane protein (gp41) permeabilize HIV-1 virions: a molecular mechanism underlying natural endogenous reverse transcription*. Proc. Natl. Acad. Sci. U. S. A., 1996. **93**(22): p. 12519-12524.
202. Rong L., Liang C., Hsu M., Kleiman L., Petitjean P., de Rocquigny H., Roques B.P., Wainberg M.A. *Roles of the human immunodeficiency virus type 1 nucleocapsid protein in annealing and initiation versus elongation in reverse transcription of viral negative-strand strong-stop DNA*. J. Virol., 1998. **72**(11): p. 9353-9358.
203. Brule F., Marquet R., Rong L., Wainberg M.A., Roques B.P., Le Grice S.F., Ehresmann B., Ehresmann C. *Structural and functional properties of the HIV-1*

- RNA-tRNA(Lys)₃ primer complex annealed by the nucleocapsid protein: comparison with the heat-annealed complex.* RNA, 2002. **8**(1): p. 8-15.
204. Druillennec S., Caneparo A., de Rocquigny H., Roques B.P. *Evidence of interactions between the nucleocapsid protein NCp7 and the reverse transcriptase of HIV-1.* J. Biol. Chem., 1999. **274**(16): p. 11283-11288.
205. Feng Y.X., Campbell S., Harvin D., Ehresmann B., Ehresmann C., Rein A. *The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site.* J. Virol., 1999. **73**(5): p. 4251-4256.
206. Shehu-Xhilaga M., Kraeusslich H.G., Pettit S., Swanstrom R., Lee J.Y., Marshall J.A., Crowe S.M., Mak J. *Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation.* J. Virol., 2001. **75**(19): p. 9156-9164.
207. Hooker C.W., Scott J., Apolloni A., Parry E., Harrich D. *Human immunodeficiency virus type 1 reverse transcription is stimulated by tat from other lentiviruses.* Virology, 2002. **300**(2): p. 226-235.
208. Khan R., Giedroc D.P. *Nucleic acid binding properties of recombinant Zn²⁺ HIV-1 nucleocapsid protein are modulated by COOH-terminal processing.* J. Biol. Chem., 1994. **269**(36): p. 22538-22546.
209. Drummond J.E., Mounts P., Gorelick R.J., Casas-Finet J.R., Bosche W.J., Henderson L.E., Waters D.J., Arthur L.O. *Wild-type and mutant HIV type 1 nucleocapsid proteins increase the proportion of long cDNA transcripts by viral reverse transcriptase.* AIDS Res. Hum. Retroviruses, 1997. **13**(7): p. 533-543.

- 210.Chang P.S., C. E. Zaia J.A., Ladne P.A., Stephens D.A., Sarver N., Rossi J.J. *Ribozyme-mediated site-specific cleavage of the HIV-1 genome*. Clin. Biotechnol., 1990. **2**, p. 23-31.
- 211.Hou E.W., Prasad R., Beard W.A., Wilson S.H. *High-level expression and purification of untagged and histidine-tagged HIV-1 reverse transcriptase*. Protein Expr. Purif., 2004. **34**(1): p. 75-86.
- 212.Sambrook J., Russell. D. W. (2001) *Molecular Cloning: A Laboratory Manual, 3rd Ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- 213.Karageorgos L., Li P., Burrell C.J. *Stepwise analysis of reverse transcription in a cell-to-cell human immunodeficiency virus infection model: kinetics and implications*. J. Gen. Virol., 1995. p.1675-1686.
- 214.Huber H.E., McCoy J.M., Seehra J.S., Richardson C.C. *Human immunodeficiency virus 1 reverse transcriptase. Template binding, processivity, strand displacement synthesis, and template switching*. J. Biol. Chem., 1989. **264**(8): p. 4669-4678.
- 215.Kati W.M., Johnson K.A., Jerva L.F., Anderson K.S. *Mechanism and fidelity of HIV reverse transcriptase*. J. Biol. Chem., 1992. **267**(36): p. 25988-25997.
- 216.Reardon J.E. *Human immunodeficiency virus reverse transcriptase. A kinetic analysis of RNA-dependent and DNA-dependent DNA polymerization*. J. Biol. Chem., 1993. **268**(12): p. 8743-8751.
- 217.Reardon J. E. *Human immunodeficiency virus reverse transcriptase: steady-state and pre-steady-state kinetics of nucleotide incorporation*. Biochemistry, 1992. **31**(18): p. 4473-4479.

218. Telesnitsky, A., and S. P. Goff. 1997. *Reverse transcriptase and the generation of retroviral DNA*. In J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), *Retroviruses*, 1997. p. 121-160. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
219. Ikeda R.A., Richardson C.C. *Enzymatic properties of a proteolytically nicked RNA polymerase of bacteriophage T7*. *J. Biol. Chem.*, 1987. **262**(8): p. 3790-3799.
220. Milligan J.F., Groebe D.R., Witherell G.W., Uhlenbeck O.C. *Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates*. *Nucleic Acids Res.*, 1987. **15**(21): p. 8783-8798.
221. Milligan J.F., Uhlenbeck O.C. *Synthesis of small RNAs using T7 RNA polymerase*. *Methods Enzymol.*, 1989. **180**: p. 51-62.
222. Kholod N., Vassilenko K., Shlyapnikov M., Ksenzenko V., Kisselev L. *Preparation of active tRNA gene transcripts devoid of 3'-extended products and dimers*. *Nucleic Acids Res.*, 1998. **26**(10): p. 2500-2501.
223. Ferre-D'Amare A.R., Doudna J.A. *Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed RNA*. *Nucleic Acids Res.*, 1996. **24**(5): p. 977-978.
224. Buiser R.G., Bambara R.A., Fay P.J. *Pausing by retroviral DNA polymerases promotes strand transfer from internal regions of RNA donor templates to homopolymeric acceptor templates*. *Biochim. Biophys. Acta.*, 1993. **1216**(1): p. 20-30.

225. DeStefano J.J., Mallaber L.M., Rodriguez-Rodriguez L., Fay P.J., Bambara R.A. *Requirements for strand transfer between internal regions of heteropolymer templates by human immunodeficiency virus reverse transcriptase.* J. Virol., 1992. **66**(11): p. 6370-6378.
226. Mely Y., de Rocquigny H., Sorinas-Jimeno M., Keith G., Roques B.P., Marquet R., Gerard D. *Binding of the HIV-1 nucleocapsid protein to the primer tRNA(3Lys), in vitro, is essentially not specific.* J. Biol. Chem., 1995. **270**(4): p. 1650-1656.
227. Tisne C., Roques B.P., Dardel F. *Heteronuclear NMR studies of the interaction of tRNA(Lys)₃ with HIV-1 nucleocapsid protein.* J. Mol. Biol., 2001. **306**(3): p. 443-454.
228. Tisne C., Roques B.P., Dardel F. *Specific recognition of primer tRNA Lys 3 by HIV-1 nucleocapsid protein: involvement of the zinc fingers and the N-terminal basic extension.* Biochimie, 2003. **85**(5): p. 557-561.
229. Tisne C., Roques B.P., Dardel F. *The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome.* J. Biol. Chem., 2004. **279**(5): p. 3588-3595.
230. Ramboarina S., Srividya N., Atkinson R.A., Morellet N., Roques B.P., Lefevre J.F., Mely Y., Kieffer B. *Effects of temperature on the dynamic behaviour of the HIV-1 nucleocapsid NCp7 and its DNA complex.* J. Mol. Biol., 2002. **316**(3): p. 611-627.
231. Stote R.H., Kellenberger E., Muller H., Bombarda E., Roques B.P., Kieffer B., Mely Y. *Structure of the His44 --> Ala single point mutant of the distal finger*

- motif of HIV-1 nucleocapsid protein: a combined NMR, molecular dynamics simulation, and fluorescence study.* Biochemistry, 2004. **43**(24): p. 7687-7697.
232. Bloomfield V.A. *DNA condensation.* Curr. Opin. Struct. Biol., 1996. **6**(3): p.334-341.
233. Heilman-Miller S.L., Thirumalai D., Woodson S.A. *Role of counterion condensation in folding of the Tetrahymena ribozyme. I. Equilibrium stabilization by cations.* J. Mol. Biol., 2001. **306**(5): p. 1157-1166.
234. Heilman-Miller S.L., Pan J., Thirumalai D., Woodson S.A. *Role of counterion condensation in folding of the Tetrahymena ribozyme. II. Counterion-dependence of folding kinetics.* J. Mol. Biol., 2001. **309**(1): p.57-68.
235. Koculi E., Lee N.K., Thirumalai D., Woodson S.A. *Folding of the Tetrahymena ribozyme by polyamines: importance of counterion valence and size.* J. Mol. Biol., 2004. **341**(1): p.27-36.
236. Tanchou V., Gabus C., Rogemond V., Darlix J.L. *Formation of stable and functional HIV-1 nucleoprotein complexes in vitro.* J. Mol. Biol., 1995. **252**(5): p. 563-571.
237. Carteau S., Gorelick R.J., Bushman F.D. *Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein.* J. Virol., 1999. **73**(8): p. 6670-6679.
238. Guo J., Wu T., Kane B.F., Johnson D.G., Henderson L.E., Gorelick R.J., Levin J.G. *Subtle alterations of the native zinc finger structures have dramatic effects*

on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. J. Virol., 2002. 76(9): p. 4370-4378.

239. Jossinet F., Paillart J.C., Westhof E., Hermann T., Skripkin E., Lodmell J.S., Ehresmann C., Ehresmann B., Marquet R. *Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. RNA, 1999. 5(9): p. 1222-1234.*
240. Skripkin E., Paillart J.C., Marquet R., Ehresmann B., Ehresmann C. *Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. Proc. Natl. Acad. Sci. U. S. A., 1994. 91(11): p. 4945-4949.*
241. Laughrea M., Jette L. *A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. Biochemistry, 1994. 33(45): p. 13464-13474.*
242. Sakuragi J, Ueda S, Iwamoto A, Shioda T. *Possible role of dimerization in human immunodeficiency virus type 1 genome RNA packaging. J. Virol., 2003. 77(7): p. 4060-4069.*
243. Paillart J.C., Marquet R., Skripkin E., Ehresmann C., Ehresmann B. *Dimerization of retroviral genomic RNAs: structural and functional implications. Biochimie, 1996. 78(7): p. 639-653.*
244. Jones J.S., Allan R.W., Temin H.M. *One retroviral RNA is sufficient for synthesis of viral DNA. J. Virol., 1994. 68(1): p. 207-216.*
245. Paillart J.C., Berthoux L., Ottmann M., Darlix J.L., Marquet R., Ehresmann B., Ehresmann C. *A dual role of the putative RNA dimerization initiation site of*

- human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis.* J. Virol., 1996. **70**(12): p. 8348-8354.
246. Abbotts J., Bebenek K., Kunkel T.A., Wilson S.H. *Mechanism of HIV-1 reverse transcriptase. Termination of processive synthesis on a natural DNA template is influenced by the sequence of the template-primer stem.* J. Biol. Chem., 1993. **268**(14): p. 10312-10323.
247. DeStefano J.J., Buiser R.G., Mallaber L.M., Fay P.J., Bambara R.A. *Parameters that influence processive synthesis and site-specific termination by human immunodeficiency virus reverse transcriptase on RNA and DNA templates.* Biochim. Biophys. Acta., 1992. **1131**(3): p. 270-280.
248. Hostomsky Z., Hostomska Z., Hudson G.O., Moomaw E.W., Nides B.R. *Reconstitution in vitro of RNase H activity by using purified N-terminal and C-terminal domains of human immunodeficiency virus type 1 reverse transcriptase.* Proc. Natl. Acad. Sci. U. S. A., 1991. **88**(4): p. 1148-1152.
249. Peliska J.A., Balasubramanian S., Giedroc D.P., Benkovic S.J. *Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity.* Biochemistry, 1994. **33**(46): p. 13817-13823.
250. DeStefano J.J., Buiser R.G., Mallaber L.M., Bambara R.A., Fay P.J. *Human immunodeficiency virus reverse transcriptase displays a partially processive 3' to 5' endonuclease activity.* J. Biol. Chem., 1991. **266**(36): p. 24295-24301.

251. Balakrishnan M., Fay P.J., Bambara R.A. *The kissing hairpin sequence promotes recombination within the HIV-1 5' leader region.* J. Biol. Chem., 2001. **276**(39): p. 36482-36492.
252. Balakrishnan M., Roques B.P., Fay P.J., Bambara R.A. *Template dimerization promotes an acceptor invasion-induced transfer mechanism during human immunodeficiency virus type 1 minus-strand synthesis.* J. Virol., 2003. **77**(8): p. 4710-4721.
253. Srinivasan A., York D., Jannoun-Nasr R., Kalyanaraman S., Swan D., Benson J., Bohan C., Luciw P.A., Schnoll S., Robinson R.A., *et al.* *Generation of hybrid human immunodeficiency virus by homologous recombination.* Proc. Natl. Acad. Sci. U. S. A., 1989. **86**(16): p. 6388-6392.
254. DeStefano J.J. *Kinetic analysis of the catalysis of strand transfer from internal regions of heteropolymeric RNA templates by human immunodeficiency virus reverse transcriptase.* J. Mol. Biol., 1994. **243**(4): p. 558-567.
255. Suo Z., Johnson K.A. *Effect of RNA secondary structure on the kinetics of DNA synthesis catalyzed by HIV-1 reverse transcriptase.* Biochemistry, 1997. **36**(41): p. 12459-12467.
256. Wu W., Henderson L.E., Copeland T.D., Gorelick R.J., Bosche W.J., Rein A., Levin J.G. *Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse transcriptase pausing at a secondary structure near the murine leukemia virus polypurine tract.* J. Virol., 1996. **70**(10): p. 7132-7142.

257. Kim J.K., Palaniappan C., Wu W., Fay P.J., Bambara R.A. *Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1*. J. Biol. Chem., 1997. **272**(27): p. 16769-16777.
258. Derebail S.S., DeStefano J.J. *Mechanistic analysis of pause site-dependent and -independent recombinogenic strand transfer from structurally diverse regions of the HIV genome*. J. Biol. Chem., 2004. **279**(46): p. 47446-47454.
259. Derebail S.S., Heath M.J., DeStefano J.J. *Evidence for the differential effects of nucleocapsid protein on strand transfer in various regions of the HIV genome*. J. Biol. Chem., 2003. **278**(18): p. 15702-15712.
260. Quinones-Mateu M.E., Gao Y., Ball S.C., Marozsan A.J., Abraha A., Arts E.J. *In vitro intersubtype recombinants of human immunodeficiency virus type 1: comparison to recent and circulating in vivo recombinant forms*. J Virol., 2002. **76**(19): p. 9600-9613.
261. Moumen A., Polomack L., Unge T., Veron M., Buc H., Negroni M. *Evidence for a mechanism of recombination during reverse transcription dependent on the structure of the acceptor RNA*. J. Biol. Chem., 2003. **278**(18): p. 15973-15982.
262. Blain S.W., Goff S.P. *Effects on DNA synthesis and translocation caused by mutations in the RNase H domain of Moloney murine leukemia virus reverse transcriptase*. J. Virol., 1995. **69**(7): p. 4440-4452.
263. Smith C.M., Smith J.S., Roth M.J. *RNase H requirements for the second strand transfer reaction of human immunodeficiency virus type 1 reverse transcription*. J. Virol., 1999. **73**(8): p. 6573-6581.

264. Gao H.Q., Boyer P.L., Arnold E., Hughes S.H. *Effects of mutations in the polymerase domain on the polymerase, RNase H and strand transfer activities of human immunodeficiency virus type 1 reverse transcriptase.* J. Mol. Biol., 1998. **277**(3): p. 559-572.
265. Cameron C.E., Ghosh M., Le Grice S.F., Benkovic S.J. *Mutations in HIV reverse transcriptase which alter RNase H activity and decrease strand transfer efficiency are suppressed by HIV nucleocapsid protein.* Proc. Natl. Acad. Sci. U. S. A. 1997. **94**(13): p. 6700-6705.
266. Miller M.D., Wang B., Bushman F.D. *Human immunodeficiency virus type 1 preintegration complexes containing discontinuous plus strands are competent to integrate in vitro.* J. Virol., 1995. **69**(6): p. 3938-3944.
267. Mansky L.M., Temin H.M. *Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase.* J. Virol., 1995. **69**(8): p. 5087-5094.
268. Jung A., Maier R., Vartanian J.P., Bocharov G., Jung V., Fischer U., Meese E., Wain-Hobson S., Meyerhans A. *Multiply infected spleen cells in HIV patients.* Nature, 2002. **418**(6894): p 144.
269. Ho D.D., Neumann A.U., Perelson A.S., Chen W., Leonard J.M., Markowitz M. *Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection.* Nature, 1995. **373**(6510): p. 123-126.
270. Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, Lifson JD, Bonhoeffer S, Nowak MA, Hahn BH, *et al.* *Viral dynamics in human immunodeficiency virus type 1 infection.* Nature, 1995. **373**(6510): p. 117-122.

271. Chun T.W., Carruth L., Finzi D., Shen X., DiGiuseppe J.A., Taylor H., Hermankova M., Chadwick K., Margolick J., Quinn T.C., Kuo Y.H., Brookmeyer R., Zeiger M.A., Barditch-Crovo P., Siliciano R.F. *Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection*. *Nature*, 1997. **387**(6629): p. 183-188.
272. Domingo E., Holland J.J. *RNA virus mutations and fitness for survival*. *Annu. Rev. Microbiol.*, 1997. **51**: p. 151-178.
273. Richman D.D., Wrinn T., Little S.J., Petropoulos C.J. *Rapid evolution of the neutralizing antibody response to HIV type 1 infection*. *Proc. Natl. Acad. Sci. U. S. A.*, 2003. **100**(7): p. 4144-4149.
274. Wei X., Decker J.M., Wang S., Hui H., Kappes J.C., Wu X., Salazar-Gonzalez J.F., Salazar M.G., Kilby J.M., Saag M.S., Komarova N.L., Nowak M.A., Hahn B.H., Kwong P.D., Shaw G.M. *Antibody neutralization and escape by HIV-1*. *Nature*, 2003. **422**(6929): p. 307-312.
275. Borrow P., Lewicki H., Wei X., Horwitz M.S., Peffer N., Meyers H., Nelson J.A., Gairin J.E., Hahn B.H., Oldstone M.B., Shaw G.M. *Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus*. *Nat. Med.*, 1997. **3**(2): p. 205-211.
276. Cao J., McNevin J., Malhotra U., McElrath M.J. *Evolution of CD8+ T cell immunity and viral escape following acute HIV-1 infection*. *J. Immunol.*, 2003. **171**(7): p. 3837-3846.

- 277.Price D.A., Goulder P.J., Klenerman P., Sewell A.K., Easterbrook P.J., Troop M., Bangham C.R., Phillips R.E. *Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection*. Proc. Natl. Acad. Sci. U. S. A., 1997. **94**(5): p. 1890-1895.
- 278.Johnson V.A., Brun-Vezinet F., Clotet B., Conway B., D'Aquila R.T., Demeter L.M., Kuritzkes D.R., Pillay D., Schapiro J.M., Telenti A., Richman D.D. *Drug resistance mutations in HIV-1*. Top. HIV. Med., 2003.**11**(6): p. 215-221.
- 279.Fisher A.G., Ensoli B., Looney D., Rose A., Gallo R.C., Saag M.S., Shaw G.M., Hahn B.H., Wong-Staal F. *Biologically diverse molecular variants within a single HIV-1 isolate*. Nature, 1988. **334**(6181): p. 444-447.
- 280.Starcich B.R., Hahn B.H., Shaw G.M., McNeely P.D., Modrow S., Wolf H., Parks E.S., Parks W.P., Josephs S.F., Gallo R.C., *et al.* *Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS*. Cell, 1986. **45**(5): p. 637-648.
- 281.Willey R.L., Rutledge R.A., Dias S., Folks T., Theodore T., Buckler C.E., Martin M.A. *Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus*. Proc. Natl. Acad. Sci. U. S. A., 1986. **83**(14): p. 5038-5042.
- 282.Coffin J.M. *Genetic variation in AIDS viruses*. Cell, 1986. **46**(1): p. 1-4.
- 283.Robertson D.L., Sharp P.M., McCutchan F.E., Hahn B.H. *Recombination in HIV-1*. Nature, 1995. **374**(6518): p.124-126.
- 284.Chen J., Dang Q., Unutmaz D., Pathak V.K., Maldarelli F., Powell D., Hu W.S. *Mechanisms of nonrandom human immunodeficiency virus type 1 infection and*

double infection: preference in virus entry is important but is not the sole factor.
J. Virol., 2005. **79**(7): p. 4140-4149.

285. Dang Q., Chen J., Unutmaz D., Coffin J.M., Pathak V.K., Powell D., KewalRamani V.N., Maldarelli F., Hu W.S. *Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways.* Proc. Natl. Acad. Sci. U. S. A., 2004. **101**(2): p. 632-637.

286. Tovanabutra S., Beyrer C., Sakkhachornphop S., Razak M.H., Ramos G.L., *et al.* *The changing molecular epidemiology of HIV type 1 among northern Thai drug users, 1999 to 2002.* AIDS Res. Hum. Retroviruses, 2004. **20**(5): p. 465-475.

287. Harris M.E., Serwadda D., Sewankambo N., Kim B., Kigozi G., Kiwanuka N., Phillips J.B., Wabwire F., Meehen M., Lutalo T., Lane J.R., Merling R., Gray R., Wawer M., Birx D.L., Robb M.L., McCutchan F.E. *Among 46 near full length HIV type 1 genome sequences from Rakai District, Uganda, subtype D and AD recombinants predominate.* AIDS Res. Hum. Retroviruses, 2002. **18**(17): p. 1281-1290.

288. Hoelscher M., Dowling W.E., Sanders-Buell E., Carr J.K., Harris M.E., Thomschke A., Robb M.L., Birx D.L., McCutchan F.E. *Detection of HIV-1 subtypes, recombinants, and dual infections in east Africa by a multi-region hybridization assay.* AIDS, 2002. **16**(15): p. 2055-2064.

289. Galetto R., Moumen A., Giacomoni V., Veron M., Charneau P., Negroni M. *The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo.* J. Biol. Chem., 2004. **279**(35): p. 36625-36632.

290. Galetto R., Giacomoni V., Veron M., Negroni M. *Dissection of a circumscribed recombination hot spot in HIV-1 after a single infectious cycle*. J. Biol. Chem., 2006. **281**(5): p. 2711-2720.
291. Besnier C., Takeuchi Y., Towers G. *Restriction of lentivirus in monkeys*. Proc. Natl. Acad. Sci. U. S. A., 2002. **99**(18): p. 11920-11925.
292. Cowan S., Hatzioannou T., Cunningham T., Muesing M.A., Gottlinger H.G., Bieniasz P.D. *Cellular inhibitors with Fv1-like activity restrict human and simian immunodeficiency virus tropism*. Proc. Natl. Acad. Sci. U. S. A., 2002. **99**(18): p. 11914-11919.
293. Bowerman B., Brown P.O., Bishop J.M., Varmus H.E. *A nucleoprotein complex mediates the integration of retroviral DNA*. Genes Dev., 1989. **3**(4): p. 469-478.
294. Farnet C.M., Haseltine W.A. *Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex*. J. Virol., 1991. **65**(4): p. 1910-1915.
295. Bukrinsky M.I., Sharova N., McDonald T.L., Pushkarskaya T., Tarpley W.G., Stevenson M. *Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection*. Proc. Natl. Acad. Sci. U. S. A., 1993. **90**(13): p. 6125-6129.
296. Schwartz O., Marechal V., Danos O., Heard J.M. *Human immunodeficiency virus type 1 Nef increases the efficiency of reverse transcription in the infected cell*. J. Virol., 1995. **69**(7): p. 4053-4059.

- 297.von Schwedler U., Song J., Aiken C., Trono D. *Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells.* J. Virol., 1993. **67**(8): p. 4945-4955.
- 298.Goncalves J., Korin Y., Zack J., Gabuzda D. *Role of Vif in human immunodeficiency virus type 1 reverse transcription.* J. Virol., 1996. **70**(12): p. 8701-8709.
- 299.Ott D.E. *Potential roles of cellular proteins in HIV-1.* Rev. Med. Virol., 2002.**12**(6): p. 359-374.
- 300.Luban J., Bossolt K.L., Franke E.K., Kalpana G.V., Goff S.P. *Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B.* Cell, 1993. **73**(6): p. 1067-1078.
- 301.Franke E.K., Yuan H.E., Luban J. *Specific incorporation of cyclophilin A into HIV-1 virions.* Nature, 1994. **372**(6504): p. 359-362.
- 302.Thali M., Bukovsky A., Kondo E., Rosenwirth B., Walsh C.T., Sodroski J., Gottlinger H.G. *Functional association of cyclophilin A with HIV-1 virions.* Nature, 1994. **372**(6504): p. 363-365.
- 303.Braaten D., Franke E.K., Luban J. *Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription.* J. Virol., 1996. **70**(6): p. 3551-3560.
- 304.Benjamin J., Ganser-Pornillos B.K., Tivol W.F., Sundquist W.I., Jensen G.J. *Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography.* J. Mol. Biol., 2005. **346**(2): p. 577-588.

305. Levy J.A. *Mysteries of HIV: challenges for therapy and prevention*. Nature, 1988. **333**(6173): p. 519-522.
306. Delwart E.L., Shpaer E.G., Louwagie J., McCutchan F.E., Grez M., Rubsamen-Waigmann H., Mullins J.I. *Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes*. Science, 1993. **262**(5137): p. 1257-1261.
307. Bello G., Casado C., Garcia S., Rodriguez C., del Romero J., Lopez-Galindez C. *Co-existence of recent and ancestral nucleotide sequences in viral quasispecies of human immunodeficiency virus type 1 patients*. J. Gen. Virol., 2004. **85**(2): p. 399-407.
308. Songok E.M., Lwembe R.M., Kibaya R., Kobayashi K., Ndembu N., Kita K., Vulule J., Oishi I., Okoth F., Kageyama S., Ichimura H. *Active generation and selection for HIV intersubtype A/D recombinant forms in a coinfecting patient in Kenya*. AIDS Res. Hum. Retroviruses, 2004. **20**(2): p. 255-258.
309. Fang G., Weiser B., Kuiken C., Philpott S.M., Rowland-Jones S., Plummer F., Kimani J., Shi B., Kaul R., Bwayo J., Anzala O., Burger H. *Recombination following superinfection by HIV-1*. AIDS, 2004. **18**(2): p. 153-159.
310. Ye Y., Si Z.H., Moore J.P., Sodroski J. *Association of structural changes in the V2 and V3 loops of the gp120 envelope glycoprotein with acquisition of neutralization resistance in a simian-human immunodeficiency virus passaged in vivo*. J. Virol., 2000. **74**(24): p. 11955-11962.