

ABSTRACT

Title of Document: NUCLEOCAPSID PROTEIN MODULATES
THE SPECIFICITY OF PLUS STRAND
PRIMING AND RECOMBINATION
PATTERNS IN HUMAN
IMMUNODEFICIENCY VIRUS

Deena Thankam Jacob, Doctor of Philosophy,
2008

Directed By: Dr. Jeffrey J. DeStefano
Associate Professor
Department of Cell Biology and Molecular
Genetics

Replication in HIV (human immunodeficiency virus) occurs through reverse transcription in which the genomic single stranded RNA is copied into double stranded DNA. This process involves two priming events namely those of the minus and plus strand DNAs. The tRNA primer required to initiate the minus strand is carried by the virus into the host cell, while the plus strand primer is generated from a region of the genomic RNA called the polypurine tract (PPT). Results in this dissertation indicate a new role for HIV nucleocapsid protein (NC) in modulating the specificity of plus strand priming. For HIV, the central and 3' (PPTs) are the major sites of plus strand initiation and other primers are rarely used. Using reconstituted in vitro assays, results showed that NC greatly reduced the efficiency of extension of non-PPT RNA primers, but not PPT. Extension assays in presence of mutant NCs

show that the helix destabilization activity of NC and its ability to block the association of RT to non-PPT primers are responsible for the preferential extension of PPT in presence of NC.

The effect of varying NC and Mg^{2+} concentrations on recombination during reverse transcription was also analyzed in this thesis. NC strongly influences the efficiency of recombination as well as the location where crossovers occurred. In contrast Mg^{2+} had a smaller effect on crossover locations. Both NC and Mg^{2+} influenced the level of pausing by RT during synthesis on RNA templates although NC's effect was more profound. At high NC concentrations, pausing was nearly eliminated even in locations with high predicted secondary structure. The results suggest that RT pausing may be limited during virus replication.

NUCLEOCAPSID PROTEIN MODULATES THE SPECIFICITY OF PLUS
STRAND PRIMING AND RECOMBINATION PATTERNS IN HUMAN
IMMUNODEFICIENCY VIRUS

By

Deena Thankam Jacob

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
2008

Advisory Committee:
Dr. Jeffrey DeStefano, Chair
Dr. James Culver
Dr. Douglas Julin
Dr. Daniel Perez
Dr. Siba Samal

© Copyright by
Deena Thankam Jacob
2008

Dedication

I dedicate this thesis to my father Jacob Thomas, who inspired me to begin this journey.

Acknowledgements

‘It takes a village to raise a child’. I have never appreciated the truth of this old African saying more. It has definitely taken a village (a global one at that!) to help me get to the end of this journey. And I feel it is fitting to acknowledge all those who have walked by my side in ways big and small.

Firstly, I would like to express my deepest gratitude to my advisor Dr. Jeffrey DeStefano. I count myself fortunate to have had the opportunity to work with him. One of the best teachers I have seen, he has the gift of reaching out to students at exactly their level of understanding. Working with Jeff has broadened my outlook on subjects ranging from science to fitness to politics to music etc. It is truly inspiring to see the love he has for science. How even with a hectic teaching and administrative schedule, he always finds time for bench work. I truly appreciate his belief and patience with me. He has taught me to do good science and more importantly to enjoy myself in the process. But the most important lessons, perhaps, are the ones he has shown by example. The humility and goodness I see in him are ideals I aspire to.

I am very grateful to my advisory committee Drs. James Culver, Douglas Julin, Daniel Perez and Siba Samal for their support and guidance. Their enthusiasm and inputs have helped me think more intensely about my work and look at it from angles I wouldn’t have otherwise. I appreciate them for making time to meet with me in spite

of their busy schedules. Most of all, I value the constructive criticism they gave me without discouraging me.

I am also very thankful to Dr. Maria Monaco who was my first mentor when I joined the program and Dr. Ann Smith, who helped me realize how much I love teaching and gave me the opportunity to improve my teaching skills.

Next I would like to thank the former lab members - Bill Bohlayer, Reshma Anthony, Nirupama Narayanan, Titilope Oduyebo and Manthan Shah. I thank them for their company and for showing me the ropes as I first began in the field of molecular biology.

I count myself lucky to be in the company of the current DeStefano lab. I have grown as a person for being with Gauri who is an amazing combination of heart and brains. I thank her for pushing me, putting her foot down with me and just being there when I needed direction in any aspect, be it cloning or cooking or culture. Through all my interactions with her, Megan has always impressed me with her capacity for information and is among the most genuine people I have ever met. I find myself refreshed each day I interact with Divya who possesses a quiet cheerfulness and am indebted to her for being there whenever I needed her (especially for graciously opening her home to me as I was completing my thesis). I thank Jeffo for his wonderful company and for bringing out the child in me (at the wrong moment at times!). I also thank Katherine who is my teacher in all things American. Her acuity

and capacity for knowledge have certainly enriched my lab experience. I have also enjoyed Elizabeth's upbeat attitude. I envy her when she talks about and her heavy work load with a smile. I thank all my lab mates once again for being who they are. I wish them the very best with their work in this lab and wherever they might choose to go in the future. I shall miss them terribly.

Next, I would like to thank my friends who made my stay here more enjoyable. To Surabhi, Nirjari, Linu, Sudeshna, Uma, Shalini, Shuchi and Sherly for all the laughter, support and inputs they gave me during the past five years. Also, my sincere thanks to Cindy and Adriana who have always been my well wishers and helped me since my first semester here.

My family has played a big role in helping me make this journey. I am grateful to them and to the members of the St. Thomas Indian Orthodox Church D.C. who welcomed me and made me feel that I belong to something bigger than myself. I would especially like to thank Thomas Varghese and family, for welcoming me into their home and hearts and being second parents to me. I would also like to thank my cousins Mathew and Anitha and their sons Chris and Ivan who went out of their way to make my stay here a better experience. My heartfelt thanks to my spiritual mentor, His Grace Mar Coorilos, whose encouragement and prayers I couldn't have done without. I am also thankful for the prayers and support of my cousins Mathai and Betsy. I have benefitted and enjoyed all the discussions we had about philosophy and life. I also want to acknowledge my sister-in-law Sara for being the cheerful optimist

that she is, and my darling nephew Noah whose pictures and videos definitely helped through my thesis writing process. I couldn't have done any of this without my mother and brother. I remember asking my mother to pray me through every crucial experiment, exam and problem I faced. I want to thank her for being my tower of strength. I stand humbled by her faith and love. I am very blessed to have an amazing brother like Ruben. He has always been my guru, knowing exactly what I need even before I know it myself. His love, support and advice are truly valued.

Finally, I'd like to thank God Almighty, for giving me this wonderful opportunity, for letting me grow and enjoy the process. I thank Him for the strength I received during times of stress. But most of all, I am thankful for the people I met and the chance I got to appreciate life and science.

Table of Contents

Dedication.....	ii
Acknowledgements.....	iii
Table of Contents.....	vii
List of Tables.....	ix
List of Figures.....	x
Chapter 1: HIV and AIDS.....	1
1.1 General Introduction.....	1
1.2 Epidemiology and classification.....	2
1.3 Disease progression during HIV-1 infection.....	5
1.4 Therapy and prevention of HIV infection.....	8
1.5 Structural Characteristics of HIV.....	11
1.6 Genetic Organization and Protein Expression in HIV.....	13
1.7 HIV-1 life-cycle in the host cell.....	16
1.9 Genome replication of HIV.....	20
1.10 Reverse Transcriptase.....	23
1.11 Nucleocapsid Protein.....	26
1.12 The Polypurine Tract.....	29
1.13 Recombination.....	32
Chapter 2: The Effect of NC on Second Strand priming in HIV.....	35
2.1 Introduction:.....	35
2.2 Materials and Methods.....	39
2.3 Results.....	47
2.3.1 NC inhibits priming with non-PPT RNA primers by HIV-1 RT.....	47
2.3.2 Internal labeling experiments indicate that all non-PPT RNA priming events are inhibited by NC:.....	51
2.3.3 NC does not dissociate intact RNA primers from the template but can destabilize smaller fragments attached to the template after RT cleavage:.....	54
2.3.4 NC inhibits priming by directly preventing extension from the 3' RNA terminus:.....	56
2.3.5 An NC mutant that shows weaker binding to nucleic acid shows reduced inhibition of RT RNA priming:.....	58
2.3.6 NC inhibits non-PPT RNA priming on a long RNA template:.....	60
2.4 Discussion.....	66
Chapter 3: Effect of NC and Mg ²⁺ on HIV Recombination.....	71
3.1 Introduction.....	71
3.1.1 Significance of recombination.....	71
3.1.2 Mechanisms of recombination.....	72
3.2 Materials and Methods.....	75
3.2.1 Materials.....	75
3.2.2 Methods.....	75
3.3 Results.....	81
3.3.1 Generation of RNA Acceptors and Donors.....	81

3.3.2 Effect of Mg ²⁺ and NC on RT synthesis.....	84
3.3.3 Frequency of hot-spots in Acceptor-Donor Set I in vitro	90
3.3.4 Frequency of hot-spots in Acceptor-Donor Set I in cell culture.....	94
3.3.5 In vitro results for frequency of hot-spots in Acceptor-Donor Set II:	98
3.4 Discussion.....	100
Chapter 4: General Discussion.....	105
Bibliography	113

List of Tables

Table 1: Protein of HIV-1 and their functions.....	14
Table 2: List of primers used with their corresponding templates.....	41
Table 3: List of primers used to generate mutations in acceptor RNAs.....	77
Table 4: Strand transfer efficiency for Acceptor-Donor Set 1 in vitro.....	89

List of Figures

Figure 1: Classification and worldwide distribution of the HIV.....	4
Figure 2: The human immunodeficiency virus.....	12
Figure 3: Organization of HIV-1 genome.....	15
Figure 4: Reverse Transcription	22
Figure 5: Structural representation of HIV-1 reverse transcriptase.....	25
Figure 6: Ribbon diagram of the secondary structure and amino acid sequence of HIV-1 nucleocapsid protein.....	28
Figure 7: Nucleocapsid protein inhibits non-PPT RNA primer extension by HIV-RT.....	49
Figure 8: Autoradiogram of primer extension by RT using internal labeling.....	52
Figure 9: NC causes dissociation of cleavage products but not uncleaved primers.....	55
Figure 10: NC inhibits extension of full-length non-PPT RNAs.....	57
Figure 11: Graphs of RNA primer extension in the presence of NC mutants.....	59
Figure 12: Schematic representation of protocol for detecting RNA primed 2 nd strand synthesis on a long RNA template in the presence and absence of NC.....	61
Figure 13: NC inhibits priming by non-PPT RNA primers generated by RT during extension on a long RNA template.....	63
Figure 14: NC inhibits non-PPT RNA primer extension over a long RNA template at 1 mM Mg ²⁺	64
Figure 15: Structures of Donor and Acceptor RNAs and a schematic representation of the Acceptor-Donor system used to study recombination in vitro.....	82
Figure 16: Extension by RT in Acceptor-Donor Set 1.....	85
Figure 17: An experiment to calculate strand transfer efficiency Acceptor-Donor Set 1.....	88
Figure 18: Frequency of transfer in zones of Acceptor-Donor Set 1 in vitro.....	93
Figure 19: Cell culture system used to study recombination and results obtained	

from it.....	97
Figure 20: Frequency of transfer in each zone of Acceptor-Donor Set 2	
in vitro.....	99

List of Abbreviations

-sssDNA	- Minus strand strong stop DNA
+sssDNA	- Plus strand strong stop DNA
AIDS	- Acquired Immunodeficiency syndrome
dNTPs	- Deoxyribonucleotide triphosphates
ds	- Double-stranded
DTT	- Dithiothreitol
EDTA	- Ethylenediaminetetraacetic acid
HIV	- Human Immunodeficiency virus
IN	- Integrase enzyme
Kb	- Kilobase
kD	- Kilo Daltons
LTR	- Long terminal repeat
Min	- Minutes
NC	- Nucleocapsid protein
nt	- Nucleotides
PBS	- Primer binding site
PCR	- Polymerase chain reaction
PNK	- Polynucleotide kinase
PPT	- Polypurine tract
PR	- Protease enzyme
RNase H	- Ribonuclease H
RT	- Reverse Transcriptase

ss - Single-stranded
T_m - Melting temperature

Chapter 1: HIV and AIDS

1.1 General Introduction

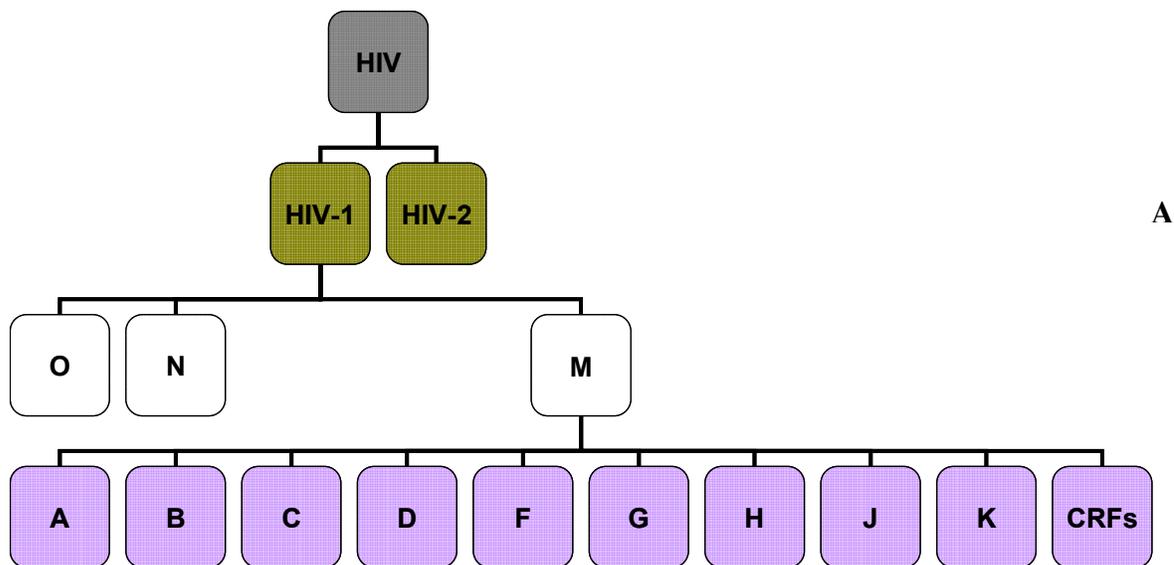
Acquired immunodeficiency syndrome (AIDS) has been an active field of study since the early 1980s. The causative agent of this disorder, HIV (human immunodeficiency virus), is a lentivirus (slow growing) of the retrovirus family. Viruses of this family replicate using a reverse flow of genetic information from RNA to DNA. The virus targets the immune system of the human host by infecting CD4+ T cells and macrophages. Lack of effective vaccines and therapy has led to AIDS becoming one of the leading causes of deaths among infectious diseases (see below). The virus was discovered in 1983 by three groups. Luc Montagnier's group at the Pasteur Institute isolated the virus from the lymph nodes of a patient and named it the lymphadenopathy associated virus (LAV) (13), while Robert Gallo and Jay Levy's groups called it the human T cell lymphotropic virus III (HTLV III) and AIDS related virus (ARV) respectively (57, 90). The universally accepted name HIV was given by the Human Retrovirus Subcommittee of the International Committee on the Taxonomy of Viruses.

1.2 Epidemiology and classification

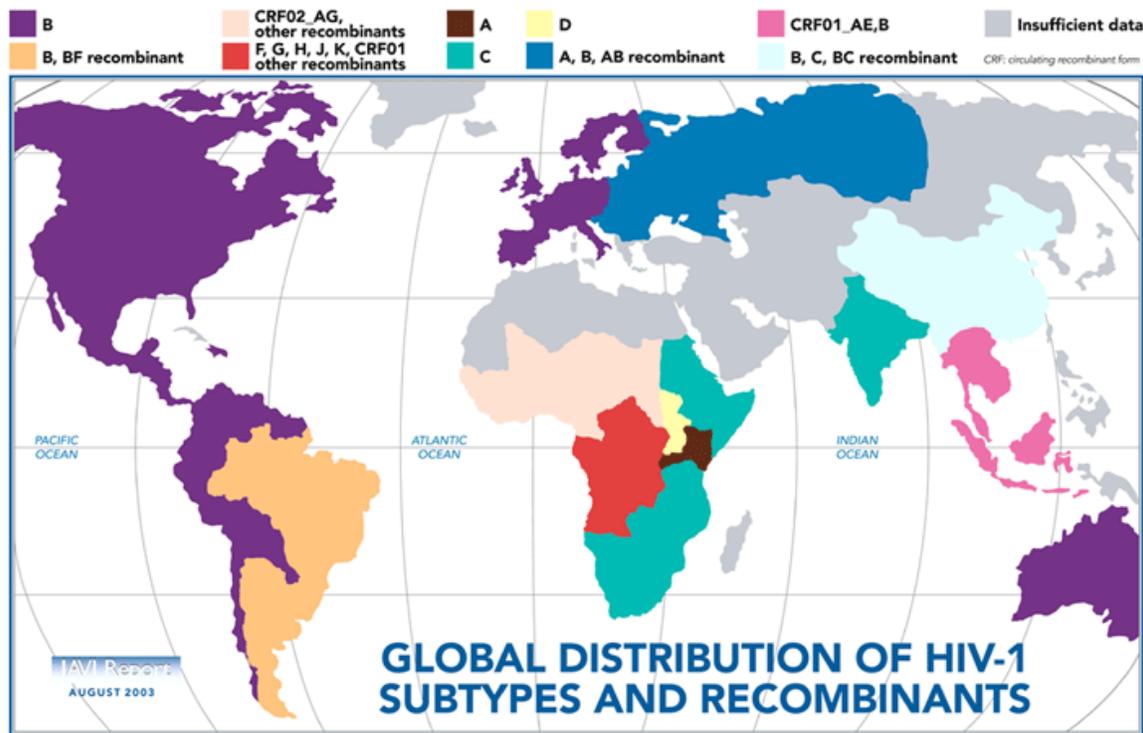
HIV, like all other retroviruses, carries its genetic information in an RNA genome. Each mature virion contains two copies of a positive sense single stranded RNA ~ 9.3 kilobases (kb) in length. Significant variation in the genome sequence between HIV isolates has led to this virus being referred to as a “quasispecies” (closely related, but nonidentical, mutant and recombinant viral genomes subjected to continuous genetic variation, competition, and selection). Based on origin and genetic differences, HIV is broadly classified into two types; HIV-1 and HIV-2. Of these, HIV-1 is the predominant cause of worldwide infections, while HIV-2 infections are concentrated in West and Central Africa and rarely found elsewhere (1). Though both HIV-1 and HIV-2 infect human beings, there is only a 60% identity between the two at the nucleotide level (65). HIV-1 is more closely related to the simian immunodeficiency virus that infects chimpanzees (SIV_{CPZ}) with almost 75% homology in the Gag protein sequence (71). Of the two, HIV-1 is more infectious and pathogenic. HIV-1 is further classified into three groups M, N and O (main, new and outlier respectively). Of the three, group M is responsible for close to 90% of HIV-1 infections in the world. This group is further divided into subtypes or clades (Fig. 1). Initial phylogenetic classification studies of HIV were based on the viral envelope sequences (93), but improvement in molecular biology and computational and phylogenetic algorithms have allowed the use of full length genomes for subtyping (107). Further adding to the diversity of HIV, recombination between different subtypes gives rise to infectious recombinant viruses termed recombinant forms

(RFs). Some of these RFs have stabilized in the population, resulting in specific circulating recombinant forms (CRFs). One particular CRF (CRF01_AE) constitutes over 3% of total HIV infections worldwide (Fig. 1 and (104)).

The earliest case now recognized as HIV in humans dates back to 1959 in Kinshasa (130). HIV is believed to have been introduced into the human population from monkeys, perhaps by human consumption. The first identified case of AIDS in the United States was in 1981. The most recent assessment shows that there are ~ 33 million people living with HIV infection in the world. In 2007 alone ~ 2.7 million new HIV infections were reported with ~ 2 million deaths due to AIDS related causes (3). The virus spreads through bodily fluids like blood and blood products, semen, vaginal secretions, breast milk etc. People at high risk include those having unprotected sex with multiple partners, people sharing needles and children born to HIV positive mothers. Also at risk are health care and laboratory workers accidentally exposed to infectious samples although such cases constitute just a small fraction of HIV infections.



A



B

Figure 1: **Classification (A) and worldwide distribution (B) of the Human immunodeficiency virus.** Adapted from www.avert.org and the IAVI Report 2003 respectively (2).

1.3 Disease progression during HIV-1 infection

HIV infection is characterized by a gradual decrease in T cell levels (specifically T cells carrying the CD4 marker on their surface which include T helper cells) typically occurring over the course of several years (61). Infection can be divided into four stages as listed below:

- i. Incubation period: This period immediately follows exposure to the virus. No symptoms are displayed and this stage may last from two to four weeks.
- ii. Acute infection: Acute infection follows the initial incubation period (~ 28 days) and lasts from 1 to 4 weeks. During this stage there is a rapid proliferation of virus accompanied by a marked drop in the level of circulating CD4+ T cells in the blood from normal levels of ~ 1200 cells/ μ L to ~ 600 cells/ μ L (61). This viremia gives rise to a CD8+ T cell (cell mediated) response as well as an antibody mediated response. The strength of the primary CD8+ T cell response has been linked to the rate of progression to AIDS (98). Together, these lower the amount of virus in the blood and cause a rise in the CD4+ T cell level to ~ 800 cells/ μ L. This response cannot, however, clear the host system of the virus completely. The symptoms, if seen at this stage, are quite non-specific and may include fever, malaise, swollen lymph nodes, pharyngitis, oral sores etc. At this stage of the infection, the patient is much more infectious.

- iii. Latency stage: The amount of virus in the blood stream is controlled over a period of time which is called the latency period. This stage could last from several months to twenty years and typically lasts ~ 8-10 years in adults. The latency period varies according to age and general health of the patient. In children and infants especially, the onset of AIDS is faster than in adults. It was initially thought that the virus goes into a quiescent mode during which the number of CD4+ T cells remains fairly high. This, however, was disproved with more sensitive assays measuring the viral RNA load in the blood stream. During the latent period ~ 10 billion viruses are produced each day. Despite this the viral load remains low and constant as the immune system is able to destroy infected cells and viruses. The CD4+ T cell balance is maintained by an accelerated replenishment rate. The virus, at this stage, is particularly active in the lymphoid organs where it gets trapped in the follicular dendritic network.
- iv. AIDS: This is the final stage of HIV infection and is used to describe a highly decimated CD4+ T cell count and the appearance of various opportunistic infections. According to the WHO, a CD4+ T cell count of less than 200/ μ L along with indicator infections signals the onset of AIDS. The opportunistic infections associated with AIDS are not restricted to the immune system. With the rapid decrease in cell mediated immunity, all organ systems become vulnerable to infection. Infections like pneumocystis pneumonia, tuberculosis (which is pulmonary during early HIV infection, but later becomes systemic) are frequently seen pulmonary diseases in AIDS patients. The most common

diseases affecting the gastrointestinal tract include fungal or viral esophagitis, chronic diarrhea of bacterial or parasitic origin. Toxoplasma encephalitis, cryptococcal meningitis, progressive multifocal leukoencephalopathy and AIDS dementia complex are neurological and psychiatric effects of AIDS. The incidence of tumors and cancers increases significantly in AIDS patients. The most common of these include Kaposi's sarcoma, Burkitt's lymphoma, and cervical cancer in HIV infected women due to the human papillomavirus.

1.4 Therapy and prevention of HIV infection

While prevention of exposure is the only effective way to control the HIV epidemic, the onset of AIDS can be delayed perhaps indefinitely in some cases, by the use of highly active anti-retroviral therapy (HAART). The first anti-retroviral drug to be approved for treatment of AIDS and HIV infection was AZT (azidothymidine) (53). This drug has especially been useful in reducing the vertical transmission of the infection from mother to child. HAART suffers from the drawback of resistance development. The virus displays a high degree of genetic variability due to an error prone replicase and extensive recombination. This gives rise to drug resistant mutants that spread through the population rendering anti-HIV drugs ineffective. To overcome this, therapy regimens using multiple classes of anti-retroviral drugs are prescribed. The major classes of anti-retroviral drugs approved for use (50) thus far are:

- Nucleoside reverse transcriptase inhibitors: Drugs in this category mimic nucleoside molecules and get incorporated into the viral DNA during reverse transcription. These drugs are typically “chain-terminators” and once incorporated they do not allow further extension thus blocking the viral polymerase. Examples include AZT, Didanosine, Abacavir etc.
- Non-nucleoside reverse transcriptase inhibitors: This class of reverse transcriptase inhibitors blocks the enzyme by directly binding close to its

active site, thus preventing polymerase activity. Drugs in this class include Etravirine, Delaviridine, Nevarapine etc.

- Protease inhibitors: The other enzyme targeted for anti-HIV therapy is protease which aids in the maturation of the virus following budding by the proteolytic cleavage of polyproteins. Such inhibitors (e.g. Indinavir, Ritonavir, Tipranavir) bind to the active site of the enzyme and prevent the release of core proteins thereby preventing maturation of released virions.
- Fusion inhibitors: Fusion or entry inhibitors are used to prevent the entry of the virus into a new host by blocking gp41 (an envelope protein) required for fusion (see section 1.7). Enfuviritide is an example of a fusion inhibitor which is approved only in cases where other drugs have proved ineffective. Another entry inhibitor Maraviroc functions by blocking the interaction between HIVgp120 and CCR5 on the host cell surface. The specific CCR5 co-receptor antagonism of this drug prevents it from being useful against CXCR4 or dual tropic HIV.
- Integrase and strand transfer inhibitors: This class of drugs targeting the third enzyme of HIV has been developed by Merck (Raltegravir). It inhibits integrase mediated strand transfer during infection thus preventing the proviral DNA from being inserted into the host genome.
- Newer approaches: Many new methods are being tested to prevent and stop HIV infection. A few of these include inhibitors of viral release, nucleic acid aptamer based inhibitors (33, 34) and siRNAs (small inhibitory RNA) (123), as well inhibitors of the viral Tat protein (29) among others.

Drug cocktails containing three drugs of two or more classes are currently used to prevent progression of the disease. However, emergence of triple drug resistant mutants has emphasized the need of developing newer and more effective drugs.

Due to the highly variable nature of the virus (see above), an effective vaccine has not yet been developed and there is no guarantee that such a vaccine can be developed in the future. The only effective weapon against the virus is to prevent its spread to more people. Educating people, especially those in the high risk groups, about HIV and its transmission is very important to this end. Practicing safe sex, avoiding needle sharing, anti-retroviral treatment of HIV positive mothers, immediate attention in the case of accidental exposure etc. are a few methods that can be employed to prevent the spread.

1.5 Structural Characteristics of HIV

HIV is classified as a retrovirus of the genus *Lentivirus*. It is a “complex” as opposed to “simple” retrovirus because it contains regulatory non-virion proteins involved in gene expression. It is spherical, enveloped and ~ 0.1 µm in diameter. Like all retroviruses, HIV is surrounded by an envelope consisting of a host-derived lipid bilayer and virus-encoded membrane glycoproteins (12). The mature virus contains external spikes of the gp120 glycoprotein and gp41 that spans the membrane (Fig. 2). The layer immediately below the external membrane is called the matrix and is composed of the ~17 kD matrix protein (MA, p17). The bullet shaped core of the virus is made of the ~24 kD capsid protein (CA, p24), that is antigenic and is predominantly used in the detection of HIV. The core also holds the genetic material of the virus consisting of two copies of a single strand positive sense RNA (hence called pseudo-diploid) coated with the viral ~ 6.5 kD nucleocapsid protein (NC, p7). The three enzymes of HIV, reverse transcriptase (RT, a dimer composed of 51 and 66 kD subunits), protease (PR, ~ 10 kD) and integrase (IN, ~ 32 kD), along with the host-derived tRNA^{Lys3} primer are also contained within the core. Several other host-derived proteins are also present in small amount within the virion or on the envelope. Though most of these are likely acquired non-specifically during budding, some may serve a role in infection (35).

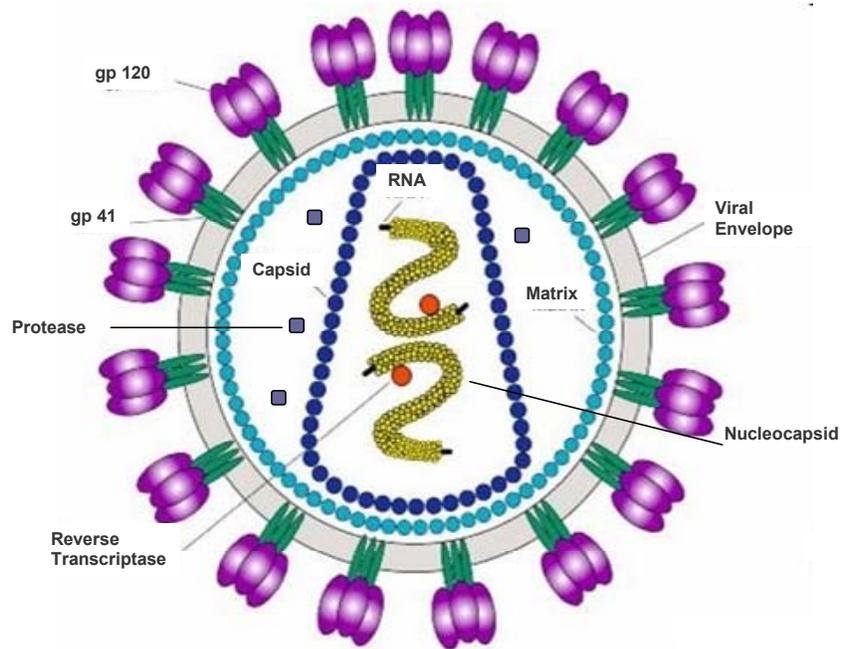


Figure 2: **The human immunodeficiency virus.** Adapted from <http://www.stanford.edu/group/virus/retro/2005gongishmail/hiv1.jpg>

1.6 Genetic Organization and Protein Expression in HIV

The genome of HIV is comprised of two copies of a single stranded RNA ~ 9.3 kb in length. The strands form a dimer that is non-covalently attached at the 5' end. The genome codes for 15 proteins most of which are generated by splicing of full-length (genomic) mRNA, which is the only functional transcript made from the virus (Fig. 3). The major products encoded by HIV are from the *gag*, *pol* and *env* regions that produce structural proteins (matrix, capsid, and nucleocapsid), enzymes (reverse transcriptase, protease, and integrase) and surface antigens, respectively. The protein products of each gene and their function in the viral life cycle are listed in Table 1.

Viral Protein	Number of copies/virion	Protein Function
Matrix, MA (p17 ^{Gag})	~5000	-targets Gag to the plasma membrane for viral assembly -Env incorporation -post-entry events
Capsid, CA (p24 ^{Gag})	~5000	-involved in core structure and assembly
Nucleocapsid, NC (p7 ^{Gag})	~5000	-genomic RNA packaging -reverse transcription -integration -RNA chaperone activity -assembly
p6 ^{Gag}	~5000	-promotes virion budding
Protease, PR	~250	-proteolytic cleavage of Gag and Gag-Pol polyproteins
Reverse Transcriptase, RT	~250	-DNA synthesis from DNA/RNA templates -RNase H cleavage
Integrase, IN	~250	-insertion of viral cDNA into host chromosome
Surface Glycoprotein, SU (gp120 ^{Env})	4-35 trimers	-host cell receptor binding -attachment and entry
Transmembrane glycoprotein, TM (gp 41 ^{Env})	4-35 trimers	-membrane fusion and entry
Virion Infectivity Factor, Vif	1-150	-APOBEC3G/F suppression
Viral Protein R, Vpr	~700	-increases post-entry infectivity -G2/M cell-cycle arrest
<i>trans</i> -Activator of Transcription, Tat	No	-activates viral transcription elongation
Regulator of Expression of Virion Proteins, Rev	No	-nuclear export of viral RNAs with introns
Viral Protein U, Vpu	No	-CD4/MHC downregulation -viral release
Negative Factor, Nef	Yes	-CD4/MHC downregulation -T-cell activation -blocks apoptosis -pathogenicity determinant

Table 1: **Proteins of HIV-1 and their functions.** Adapted from Swanson and Malim Cell 2008, Snapshot: HIV-1 Proteins, **133**: 742

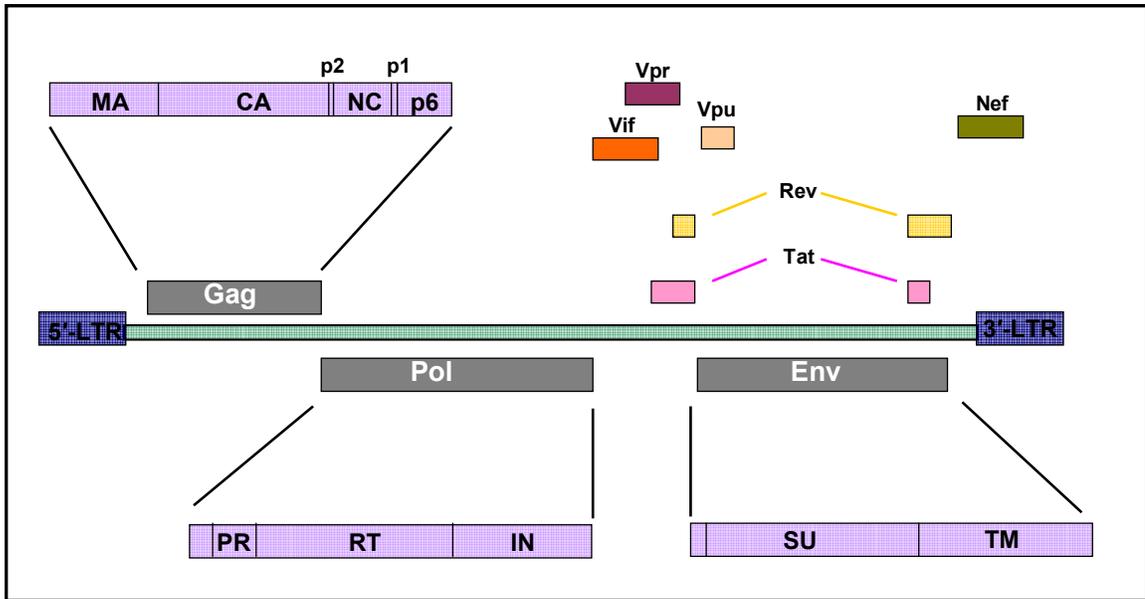


Figure 3: **Organization of HIV-1 genome.** Functions of the various gene products are listed in Table 1. Adapted from Swanson and Malim Cell 2008, Snapshot: HIV-1 Proteins, **133**: 742

1.7 HIV-1 life-cycle in the host cell

- **Entry:** HIV can enter those cells that present CD4 receptors on their surface. This receptor is usually found on immune cells such as T cells and macrophages thus making them the usual targets for HIV infection (other targets include natural killer cells, dendritic cells etc). The gp120 surface protein interacts with the CD4 marker resulting in attachment of the virus to the host cell. However, for entry, the virus also needs a co-receptor in addition to the CD4 receptor. The most commonly used co-receptors for HIV entry are CCR5 (R5) and CXCR4 (X4). The R5 ligand is usually found on macrophages while the X4 moiety is characteristic of T cells. The interaction between the CD4-gp120 complex and the co-receptor is required for changes in the cell membrane to facilitate viral entry (26).
- **Reverse Transcription:** Once the virus enters the cell, it uncoats and releases the core into the cytoplasm. The process of reverse transcription occurs in the core and involves the copying of a single stranded RNA into double stranded DNA. The first strand to be made is the minus sense DNA strand which is then used as a template for the plus sense DNA strand (discussed in more detail in the next section). The plus strand contains LTRs (long terminal repeats) at each end and a central flap region that facilitates the transport of the newly made double stranded viral DNA into the host nucleus for integration (26).

- **Integration:** The enzyme integrase facilitates the incorporation of the viral DNA into the genome of the host cell. The first step in this reaction involves the processing of the viral DNA by integrase where two nucleotides are removed from the 3' ends of both strands. This exposes the 3'-hydroxyl group on the terminal CA dinucleotide that is conserved among retroviruses and many transposons. The processed ends are then inserted into the target DNA sequence. In case of HIV-1, the sites for insertion of these ends are 5 bases apart. Once integration occurs, host cell machinery carries out gap filling and ligation thus incorporating the viral DNA into the host genome. In HIV-1, integration can essentially occur at any location on the host genome, but shows a preference for sites of active gene transcription (27).
- **Synthesis and processing of viral RNA:** LTRs generated at the ends of the viral DNA act as promoters for transcription of HIV-1 by the host RNA polymerase II. Retroviral RNA transcripts are subject to the same processing steps as cellular mRNAs including 5' capping, 3' polyadenylation followed by splicing etc. Full length RNAs are used either as the new viral genome or to encode the Gag-Pol polyprotein, while singly or multiply spliced RNAs are used to code for the rest of the viral gene products. The relative levels of spliced to full length transcripts are controlled by the virus and hence present a potential target for control of viral replication (26). Regulatory proteins of HIV can modulate the production of subgenomic RNAs e.g the buildup of Rev protein can cause an increase in the nuclear export of un-spliced singly spliced mRNAs thus leading to an increase in the production of full length

(genomic) transcripts and those coding for envelope protein (100) while Tat up-regulates transcription from the LTR leading to a general increase in viral gene products (26).

- Protein synthesis: After synthesis, full length and spliced viral RNAs are transported out of the nucleus and used to produce proteins using the translational machinery of the host cell. Proteins derived from *gag*, *pol*, and *env* are translated as polyproteins that are then processed by proteases. While Gag and Pol are processed by the viral protease (PR), Env precursor protein gp160 is cleaved by cellular proteases such as furins to produce gp120 and gp41. Full length genomic RNA codes for the Gag precursor, p55, which is proteolytically cleaved to yield MA (p17), CA (p24), NC (p7), p6 and spacer peptides p2 and p1. The full length mRNA is also used to make the Gag-Pol precursor which is cleaved to yield PR, RT, and IN, as well as the Gag proteins. The ratio between the Gag and the Gag-Pol precursors is maintained by a -1 frameshifting event which occurs due to the presence of a slippery sequence and allows read through of a stop codon at the end of the *gag* gene. Read through occurs ~ 5% of the time leading to an ~ 20:1 ratio of Gag:Gag-Pol proteins. Presumably maintenance of this ratio is important to the virus as altered ratios inhibit RNA dimerization and virion infectivity (115).
- Assembly, maturation and budding: These are the final stages in the production of new virions. During assembly, structural and enzymatic components required in a new virus are directed toward the cell surface of the host. Assembly usually occurs at discrete domains of the plasma membrane

called rafts. Gag/Gag-Pol associates with viral genomic RNA dimers via the NC portion of the precursor and the localization of Env proteins on rafts increases the affinity of Gag and associated RNA for these domains, thus transporting viral genomic RNA to the site of assembly. Multimerization of Gag causes a curvature at the location. An increase in this curvature finally leads to budding and release of the new virion. Host proteins that are required for this process have been identified and thus represent potential therapeutic targets (35). The process of protein maturation in the virion is carried out by PR and involves the cleavage of polyproteins into their active forms. This may begin during the late stages of assembly or immediately after viral release. The mature virus is then ready to infect the next host cell.

1.9 Genome replication of HIV

Genomic replication in HIV involves the reverse flow of genetic information from RNA to DNA and is catalyzed by reverse transcriptase. The copying of the single stranded plus sense RNA into double stranded DNA involves the following steps (26) (Fig. 4):

- The first strand to be synthesized in HIV is the minus strand DNA. The initiation of this step occurs using a tRNA primer that the virus carries with it from the previous host cell. This primer binds to the primer binding site (PBS) of the RNA genome ~ 182 nucleotides downstream from the 5' end. RT carries out the synthesis of the minus strand DNA up to the 5' end of the template strand while simultaneously degrading the template RNA with its RNase H activity (activity that degrades RNA that is part of an RNA-DNA hybrid, see below under. Section 1.10). The newly made DNA is ~ 164 bases long and is called the minus sense strong stop DNA (-sssDNA).
- The RNase H degradation of the template frees the –sssDNA to transfer to the 3' end of either the same template or the co-packaged RNA template. This is possible because the HIV RNA has direct repeat regions (R) at each end of the genome. This is the first obligatory strand transfer event during HIV genome replication and is facilitated by NC.
- Once the –sssDNA transfers to the 3' end of the RNA template, minus strand synthesis proceeds up to the PBS site and the template strand is almost completely degraded during this process.

- Two regions of the RNA genome, however, are resistant to this degradation and are hence left attached to the newly made minus strand. Both of these RNAs are identical 15 mers containing only purines. Based on their location, they are called the central and 3' polypurine tracts (cPPT and 3' PPT respectively).
- The PPTs are then used to initiate the synthesis of the plus strand DNA. The 3' PPT is elongated through the first 18 nts of the tRNA primer still attached to the -sssDNA. The 19th nt from the 3' end of the tRNA contains a modified base where RT terminates. This gives rise to the plus sense strong stop DNA (+sss DNA). At this point, the PPTs and the tRNA are removed by RT. The removal of the tRNA primer, frees the +sssDNA at the 3' end.
- The exposed +sssDNA is now free to carry out the second obligatory strand transfer to the PBS region of the minus strand. This presumably results in a circularized structure.
- Both strands are now able to go to completion by using each other as template by means of the strand displacement activity of RT.
- The completed product of HIV genome replication is a double stranded DNA with long terminal repeats (LTRs) at both ends (26).

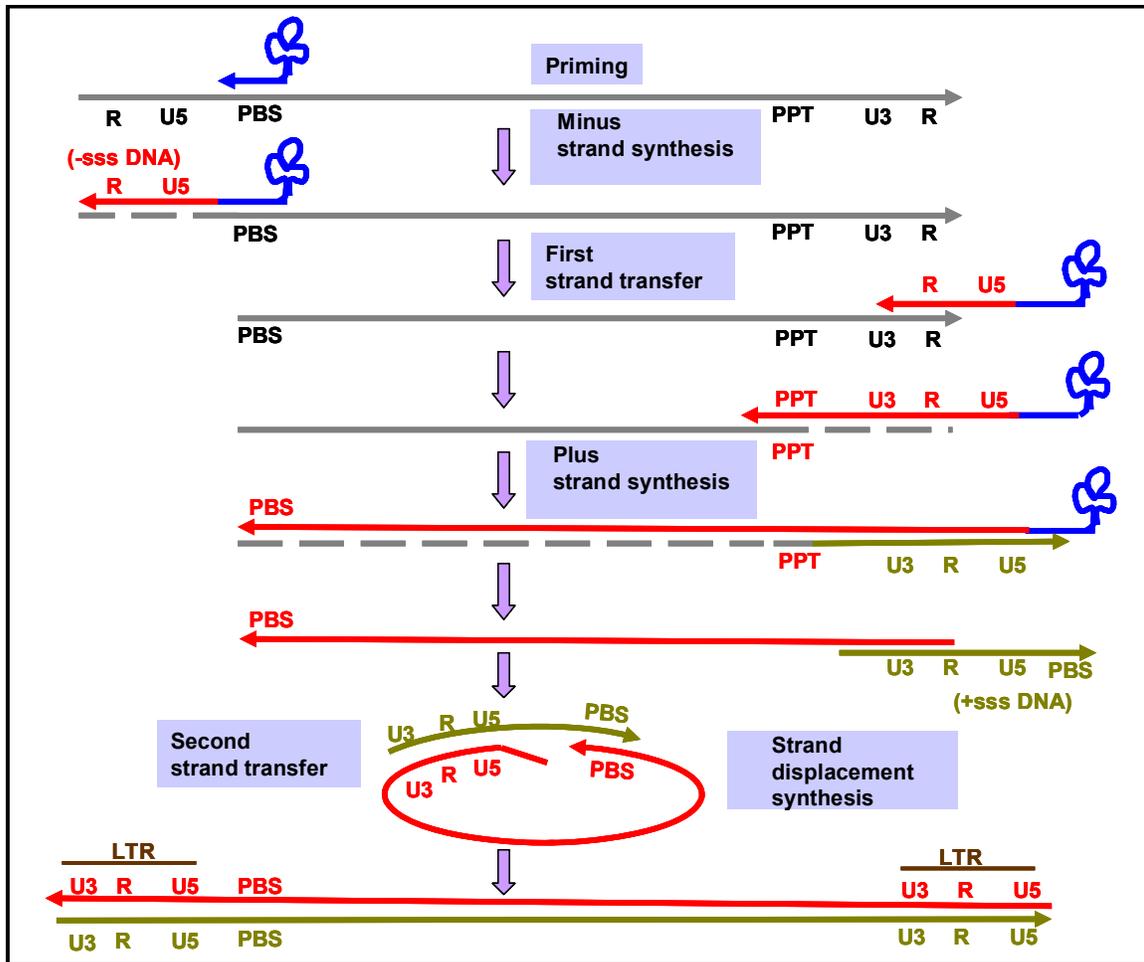


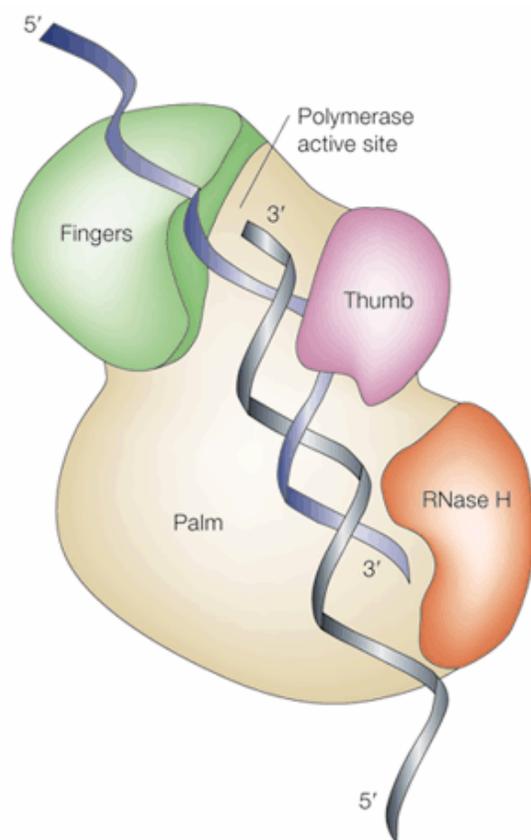
Figure 4: **Reverse Transcription.** The virus uses its polymerase to copy the genomic RNA strand into a double stranded DNA with LTRs to facilitate integration into the host genome. Adapted from Basu et al., *Virus. Res.* 2008 (14). The steps involve priming of the minus strand DNA by a tRNA primer followed by transfer of the nascent DNA strand to the 3' end of the genome. This is followed by synthesis of the minus strand and simultaneous degradation of the RNA template leading to the generation of the PPT. Plus strand DNA synthesis begins from the PPT and is completed by strand displacement synthesis leading to the completion of the double stranded DNA (see text for details).

1.10 Reverse Transcriptase

The enzyme reverse transcriptase (RT) was first discovered in the early 1970s by Howard Temin and David Baltimore (11, 122). The discovery of this enzyme upset the then widely held “central dogma” where genetic information was thought to flow only from DNA to RNA to proteins. This property of reverse transcription of genetic information is a characteristic of members of the family *Retroviridae*.

Reverse transcriptase in HIV is derived from the *pol* region of the genome. The functional enzyme consists of two subunits (66 kD and 51 kD respectively) (46). The 51 kD subunit (p51) is a product of proteolytic cleavage of the 66 kD (p66, 560 amino acids) subunit. Both subunits share a common amino terminal end while the smaller subunit lacks the 120 carboxy terminal amino acid of p66. RT is easily expressed in bacterial expression systems, which has facilitated the study of this enzyme. Each subunit can be expressed separately and then mixed in equimolar concentrations to give the functional heterodimer. The RT heterodimer possesses three functions: a) DNA-dependent DNA polymerization, b) RNA-dependent DNA polymerization, and c) Ribonuclease H (RNase H) cleavage. The latter function cleaves RNA that is part of an RNA-DNA hybrid. RT is a highly error prone replicase and like the replicases of other RNA viruses, has no proof reading activity. There are several estimates of the error rate of RT which vary significantly depending on the system used and whether the experiment was done in cells or with purified enzymes. A rate between $3-10 \times 10^{-5}$ (1 error per 10,000-33,000 nucleotides incorporated) is generally accepted (15).

The polymerase activity of the enzyme maps to the amino terminal domain while the RNase H function is located at its carboxy terminal and is absent in p51. Though the p66 subunit contains all the information needed for activity, the dimerization of RT is necessary for optimal activity. The p51 subunit has been shown to have little activity by itself and is enzymatically inactive in the p66/p51 dimer (58). RT has been shown to require divalent cations such as Mg^{2+} for optimal activity (117). There are three catalytic residues in the polymerase active site (Asp110, Asp185 and Asp186) that play a role in the metal ion binding (68). Crystallography studies show that the enzyme assumes an arrangement analogous to a right hand that is open while “grasping” the primer-template hybrid and closed otherwise. Both subunits can be divided into fingers, thumb, palm and connection domains which are arranged differently in each (68). The polymerase and the RNase H domains of RT are separated by a distance of 18 base pairs with the site of RNase H cleavage 18 bases behind the DNA primer 3' terminal nucleotide on a double stranded substrate.



Nature Reviews | Molecular Cell Biology

Figure 5: **Structural representation of HIV-1 reverse transcriptase (RT) and its various domains.** (Negroni M. and Buc H., Nature Reviews Molecular Cell Biology 2001, (95))

1.11 Nucleocapsid Protein

The nucleocapsid protein (NC) of HIV is a small, highly basic protein containing 55 amino acids. It has two 14 amino acid zinc fingers containing CCHC motifs for metal binding joined by a linker domain (Fig. 6). During maturation, NC is generated by the proteolytic cleavage of the Gag precursor. NC plays a role in assembly as part of the Gag precursor even before maturation. It has been shown to participate in several processes in the viral life cycle including packaging (6, 7, 31, 32, 106, 119), dimerization of genomic RNAs (9, 52, 84, 116), tRNA binding (as part of the Gag precursor) (17, 21, 51, 77, 109), strand transfer and recombination (5, 37, 38, 62, 63, 76, 102, 108, 129), stimulating and modulating RNase H activity of RT (24, 126), integration (18-20, 83) and RT-directed DNA synthesis (75, 103). As a chaperone protein, NC facilitates the rearrangement of nucleic acid molecules into thermodynamically more favorable structures. Two properties of NC facilitate its function as a chaperone, namely its helix destabilizing (unwinding) and nucleic acid aggregation (8, 48, 85, 87, 118, 120) activities (for a review see (88)). The former, in addition to helping RT traverse the genome by melting secondary structures, may also facilitate the removal of RNA fragments generated by RNase H that remain bound to the nascent minus sense DNA. Aggregation activity of NC may play a role in bringing nucleic acids in closer contact with each other, thereby aiding strand transfer. Aggregation is associated with the N terminal region of NC, while the zinc fingers have been shown to be important in unwinding. Unlike helicases, NC-mediated unwinding is NTP-independent and relatively weak. The ability to perform

helix destabilization is distributed unequally among the zinc fingers. The N terminal zinc finger is more critical for this function than the C terminal one. However, both zinc fingers in their appropriate locations are necessary for optimal unwinding activity (66).

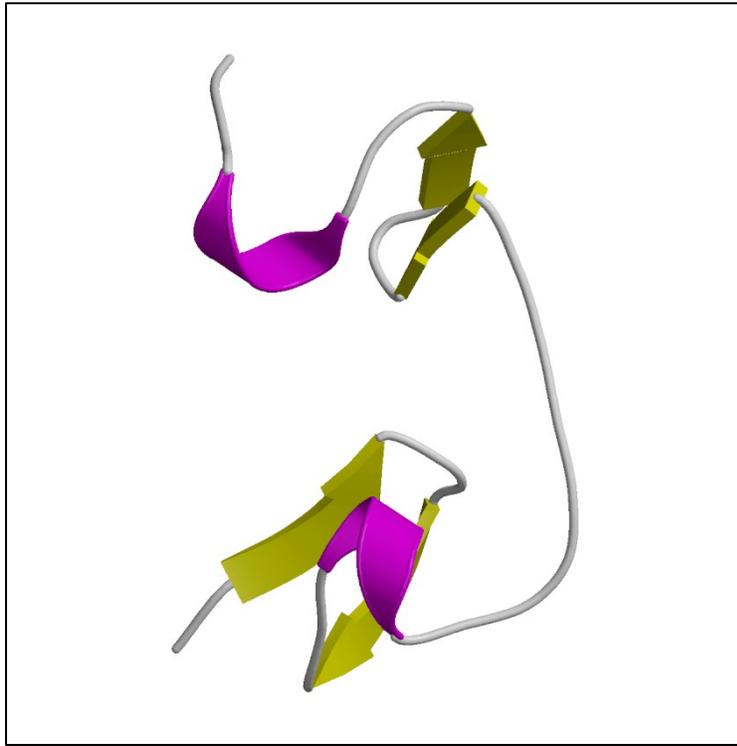
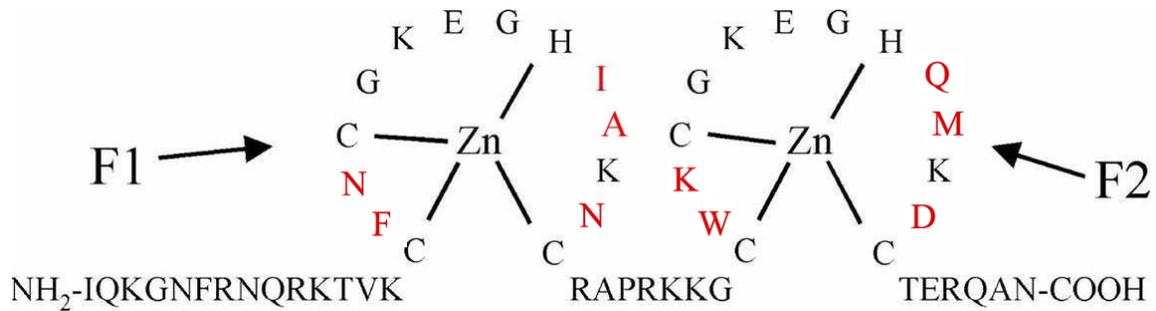
A**B**

Figure 6: **Ribbon diagram of the secondary structure (A) and amino acid sequence (B) of HIV-1 nucleocapsid protein.** The two zinc fingers are shown as F1 and F2 and the amino acid sequence differences between them are indicated in red. Ribbon diagram is taken from the Protein Structure Classification database: <http://www.cathdb.info/cathnode/4.10.60>

1.12 The Polypurine Tract

During the process of reverse transcription two RNA priming events are necessary, one for the minus strand and the other for the plus strand (see Fig. 4). In HIV, the minus strand is primed by tRNA^{Lys3} that the virus carries with it as it enters a new cell. The initiation of the plus strand, however takes place with a primer that is generated during the synthesis of the minus strand. During minus strand extension, the template RNA is degraded by the RNase H activity of RT. However, a 15 nucleotide (nt) RNA fragment is not degraded and remains attached to the template. This RNA, which is made up entirely of purines (5'-AAAAGAAAAGGGGGG) and hence called the polypurine tract (PPT), is used to initiate the synthesis of plus strand DNA. All retroviruses and retrotransposons possess polypurine tracts the sequence of which is highly conserved among retroviruses. Priming of HIV-1 plus strand DNA is restricted almost exclusively to the PPT though theoretically it is possible for RT to use other RNA fragments generated by its RNase H activity as primers. Many studies have addressed why the PPT is resistant to degradation by the RNase H activity of RT and is the preferred primer. Two aspects of PPT, its sequence and structure, have been examined in this regard. The base sequence of the PPT is highly conserved within groups of lentiviruses and very few changes are seen even between different groups (105). This suggests that the base sequence is an important player in PPT's role as a primer. Secondly, changing the context in which the PPT is placed does not alter its priming capability. For example, when the run of U's flanking the 5' end of the 3' PPT was replaced, no alteration was seen in the efficiency of priming (101). Even inserting the entire PPT into different regions of the genome did not alter its

priming capacity as initiation of the plus strand occurred from wherever PPT was placed (101). Studies have shown that within the PPT itself, changing various bases affected priming differently. Bases to the 5' end of the PPT were not found to be as crucial as bases at the 3' end (91). Mutation studies have shown that the run of 6 Gs (G tract) at the 3' end of the PPT is very critical for recognition and priming while the upstream bases are more dispensable. Even when all the bases except for the G tract were replaced, proper cleavage was seen to occur in primers in vitro (101).

Structural studies have shown that RT and PPT interact very closely. Mutations in RT have been shown to affect the cleavage and extension of PPT (105). The RNA:DNA hybrid of PPT with minus strand DNA assumes a structure different from regular DNA:DNA hybrids (78). This is because RNA helices usually take on an A-form structure while DNA helices model the B-form and together the PPT and the newly formed negative DNA strand form a helical arrangement between A and B forms (101). The size of the minor groove here is larger than A form but smaller than B form while the size of the major groove is similar to that found in B form helices. Another factor contributing to the special structure of the PPT:DNA hybrid is the fact that all the PPT bases are purines. So the hybrid has only purines on the RNA strand and only pyrimidines on the DNA strand. Studies have shown that helices with such a combination form structures of a different class (101). Also, a 13° bend has been observed at the PPT and U3 junction (3' end of PPT) which deviates from Watson-Crick geometry (81, 105). It has been suggested that this may contribute to the resistance that PPT displays towards cleavage and the specificity of cleavage at the PPT/U3 junction (81, 105). The correct excision of PPT after plus strand synthesis is

important as the PPT/U3 junction defines the start of the LTR formed after reverse transcription. The LTR not only acts as promoter for viral transcription, but is also necessary to generate the correct ends for integration into the host genome (26).

1.13 Recombination

One of the characteristic features of HIV is an extensive amount of genetic variability resulting from elevated polymerase error and recombination rates. Recombination in HIV is different from that occurring in bacterial and mammalian cells in that it occurs by template switching (from a “donor” to an acceptor” template) during reverse transcription of the viral genome. This is made possible due to the presence of two copies of the RNA genome in the virus. Recombination between two identical RNA molecules would result in progeny that are identical to the parent virus (excluding base misincorporations by RT). However, if the two strands of RNA co-packaged in a virus are different, either due to dissimilar errors in the previous round of replication or dual infections (two retroviruses infecting the same cell and both producing proviruses), the resulting progeny would differ from the virus initiating the infection. HIV displays such a high degree of genetic variability that even in cases where patients are exposed to the virus only once, there is a marked difference between the genomes of the initially infecting virus and those produced later during the course of infection. To add to this, HIV has many types, subtypes and sub-subtypes, some of which are capable of co-infection. The geographical distribution of HIV makes it possible for viruses of two clades to infect the same person resulting in the generation of inter sub-type recombinants. For example, the presence of subtypes A, B and AB recombinants in Eastern Europe. Once recombinants survive and are able to stably establish infections, they are called circulating recombinant forms (CRFs) e.g. CRF02_AG, is responsible for around 5% of total HIV infections (104).

Taking into account the mechanism of HIV replication, recombination confers genetic advantages to the virus. Having two copies of the genome enables the completion of replication even when one of the genomes is broken or damaged. Also of great value is the fact that recombination can be used by the virus to acquire evolutionary benefits like drug resistance etc. As described in the previous subsection, HIV replication requires at least two events of recombination or strand transfer (also called strand jumping or template switching). The first mandatory strand jump occurs when extension from the tRNA^{Lys3} primer reaches the 5' end of the genome. The newly made DNA (-sssDNA), jumps to the 3' end of the genome. This step requires the presence of NC. The second obligatory strand transfer event occurs when the plus strand DNA initiated from the 3' PPT reaches the tRNA^{Lys3} primer and crosses over to PBS at the 3' end of the minus strand DNA. Since both these transfer events occur after extension takes place to the end of the template, they are terminal transfers.

In addition to these compulsory transfers, internal transfers can occur at essentially any point during reverse transcription. The number of internal transfers is much higher during the generation of the minus strand than the plus strand. Minus strand recombination is RNase H-dependent and occurs by a mechanism called “copy-choice” and involves switching of the nascent DNA strand between the two copies of genomic RNA. The rate of recombination differs in different cell types. In T cells approximately 3-10 crossovers per replication cycle have been reported (131), while in macrophages, a much higher rate in the order of 30 events per virus replication cycle have been reported (89).

Though recombination is a powerful vehicle for HIV evolution and can occur between essentially any co-packaged genomes, several restrictions to the process exist. The most obvious of these is the suppression of co-infection. HIV infection causes the down regulation of CD4 markers on the cell surface thus making the attachment and entry of another virus less probable. For recombination to occur, the packaging of two different genomic RNAs is required. This places another constraint on the process as dimer signals of different subtypes may hamper co-packaging e.g. subtypes B and C do not form heterodimers due to conflicting dimer signals (25). Sequence divergence outside the dimer signal region also plays a role in the propensity to form heterodimers as shown by the strict restriction of co-packaging of HIV-1 and HIV-2 (47). Another issue with dually infected cell is the possibility of defective assembly due to protein incompatibility (82). Also, recombination within a single gene may result in a chimeric protein that may be non-functional.

The value of recombination to the replication and evolution of HIV makes it a topic of immense significance. Not only is the study of recombinants and recombination crucial for epidemiological purposes, but also to understand the evolutionary and prophylactic prospects for AIDS. The most obvious applications of recombinant studies would be to develop therapeutics that the virus is less likely to acquire resistance to, and to design vaccines bearing in mind the genetic diversity and evolutionary possibilities of the virus.

Chapter 2: The Effect of NC on Second Strand priming in HIV

2.1 Introduction:

Human immunodeficiency virus (HIV)-1 is a retrovirus with two copies of ~9.3 kb positive sense single stranded genomic RNA. Replication in HIV-1 involves copying of the RNA into double stranded DNA. The first strand to be synthesized is the minus sense DNA strand. Initiation of this strand occurs using a host tRNA as primer (tRNA^{Lys3}). Reverse transcriptase (RT), the replicative enzyme of HIV, initiates DNA synthesis ~182 nucleotides downstream of the 5' end of the RNA genome at the primer binding site (PBS) where the 3' end of the tRNA binds. Both the polymerization and RNase H activities of RT are used during reverse transcription. RT copies the genome up to the 5' end (giving rise to the minus sense strong stop DNA (-sssDNA)) and then the complex jumps to a complementary region at the 3' end of the genome and the synthesis of the minus strand DNA continues with simultaneous degradation of the template RNA. Initiation of the plus strand occurs at a site called the polypurine tract (PPT, 5'-AAAAGAAAAGGGGGG) which is a part of the RNA genome and is not degraded by RT during minus strand synthesis (26). There are two sites of plus strand initiation in the viral genome. The first is the 3' PPT found immediately adjacent to the U3 region of the viral genome and the second is found in the integrase coding region of the *pol* fragment and is called the central PPT (22, 23, 105). The specific and predominant use of the PPT by HIV is remarkable given that the two PPT regions constitute of only ~0.3% of the entire genome and RT

is capable of using non-PPT RNAs as primers (78, 101). After initiation at the 3' PPT, the plus strand is elongated until a portion of the tRNA is copied giving rise to the plus sense strong stop DNA (+sssDNA). Completion of plus strand synthesis requires another strand jump by the +sssDNA to the 3' end of the newly formed minus strand DNA. The completed double stranded DNA includes long terminal repeats (LTRs) that are used during integration into the host genome.

Though RT degrades the template during minus strand synthesis, the size of fragments generated varies with some being long enough to remain bound to the nascent minus strand DNA (44). Any of these fragments could potentially be used to prime plus strand synthesis. Retroviruses, including HIV, have been shown to use primers other than the PPT to initiate plus strand synthesis (78, 92). In cases like avian sarcoma virus (ASV), multiple sites for plus strand initiation have been reported with the completed plus strand containing numerous discontinuities (79). Although the extent of non-PPT primer usage in HIV is believed to be relatively low, the PPT is not absolutely required for replication as HIV can replicate without a PPT sequence, albeit very inefficiently (91).

Reverse transcriptase extends DNA primers more efficiently than RNA primers. Even the PPT, though highly efficient in comparison to other RNAs, is relatively poor as PPT DNA is a better primer (54). On DNA primers, RT binds with its polymerase domain oriented toward the primer 3' end which facilitates nucleotide addition. In contrast, RT binds RNA primers at the 5' end in an orientation that promotes RNase H cleavage rather than extension. However, some binding near the 3' end must occur since RNAs can be inefficiently extended (39, 42, 43, 78, 97, 101,

113). Factors that promote extension (i.e. sequence preference, structure, nucleotide length) of non-PPT RNAs are as yet unclear although studies of the PPT suggest that structure and sequence may be important (see below).

As described earlier, the sequence and structure of PPT have been implicated in its choice as the primer for plus strand synthesis (see Chapter 1). The presence of 6 G residues at the 3' end of PPT and the unique structure it forms with RT have been shown to be involved in making it a primer of choice. The 13° bend at the PPT/U3 junction may be involved in directing RT for the correct excision of the primer after plus strand synthesis (81, 105). It has also been proposed that the distinctive curvature of the PPT-DNA hybrid could aid RT binding (81). Crystal structures show that when bound to RT, a ~ 45° bend is observed in the primer-template (74, 111). The PPT structure could help RT bind more tightly if the induced curvature favored a better fit to the active site. This may also be the case for DNA-DNA hybrids as HIV-RT shows strong preferential binding to DNA oligonucleotides with sequences similar to the PPT (41).

The nucleocapsid protein (NC) of HIV acts as a nucleic acid chaperone in the virus. NC uses its helix destabilization and aggregation activities to melt secondary structures and form thermodynamically more favorable structures. NC has been shown to be involved in almost all steps of reverse transcription (for a review see (88)). The unwinding activity of NC, in addition to helping RT traverse the genome by melting secondary structures, may also facilitate the removal of RNA fragments that remain bound to the nascent minus sense DNA. This could be a mechanism to prevent spurious RNA priming.

Results presented here show that in addition to helping dissociate small RNA fragments, NC also directly inhibits RT extension of non-PPT RNAs while having no effect on PPT usage. Therefore, NC may play an important role in specifying the use of the PPT regions for second strand priming in HIV.

The aim of this project was to analyze if factors other than the resistance of PPT to internal cleavage and the orientation and affinity of RT while binding to the PPT play a role in the selection of PPT as a primer for second strand synthesis.

2.2 Materials and Methods

2.2.1 Materials:

The expression clone for wild type HIV-RT (strain HXB2) was kindly provided by Dr. Samuel Wilson (NIEHS, Research Triangle Park, NC). This enzyme (non-histidine tagged version) was purified as described (67). The RNase H deficient mutant of RT (E478Q) was kindly provided by Dr. Stuart Le Grice (NCI, Frederick, MD). The expression clone for wild type NC was a gift from Dr Charles McHenry (University of Colorado) and was purified as described (128). All mutant NCs were kind gifts from Dr. Robert Gorelick (NCI, Frederick, MD). Both NC and RT proteins were tested for purity using SDS-PAGE gels (data not shown). *Pfu* polymerase used for PCR was from Strategene while the Phusion HF kit from New England Biolabs was used to join the PCR fragments. The Rapid Ligation kit and DNA *Hinf*I digested Φ X174 ladder were from Promega. Plasmid pNL4-3 was from the NIH AIDS Research and Reference Reagent Program while pBS Δ *Pvu*II (45) was derived from pBSM13+ (Strategene). All restriction enzymes and T4 polynucleotide kinase were from New England Biolabs. Proteinase K was from Shelton Scientific Inc., CT. NP-40 was obtained from Calbiochem, D-Biotin was purchased from AnaSpec Inc., and streptavidin agarose was obtained from Pierce. GC5 *Escherichia coli* competent cells from Gene Choice were used for transformation. Plasmid preparation for sequencing was carried out using the Qiagen Miniprep kit. Radiolabel was purchased from Amersham and Perkin Elmer. Sephadex G-25 spin columns were from ISC Bioexpress. Cellulose nitrate filter discs were obtained from Whatman Inc. Non-radioactive nucleotides and SP6 RNA polymerase were from Roche Diagnostics.

Immobilized streptavidin slurry was from Pierce. All oligonucleotides were obtained from Integrated DNA Technologies and all other reagents were from Fisher and VWR.

2.2.2 Methods:

5' end-labeling of primers: Fifty pmoles of primer were 5' end labeled with γ - ^{32}P ATP using T4 polynucleotide kinase following the manufacturer's protocol. Unincorporated ATP was removed using G-25 sephadex spin columns. Labeled primers were stored at -20°C .

Hybridization of nucleic acids: Primers (5' end labeled with ^{32}P) and corresponding 55 nt templates (see Table 2) were hybridized in 80 mM KCl, 1 mM DTT, and 50 mM Tris-HCl (pH=8.0) at a ratio of 1.5:1 (primer:template) in a volume of 10-20 μl s by heating to 80°C for 3 min then slow cooling at a rate of 2° per min to 30°C . Hybridized material was used immediately in assays. For assays with the 430 nt RNA template (see below) the ratio of primer:template was 3:1. A lower ratio was employed with the small templates since hybridization would presumably be more efficient and a vast excess of labeled primer could hamper analysis of cleavage and extension products.

Extension assay with 55 nt templates: Primer-template (6 nM final in 55 nt template) was incubated with or without of NC (2 μM final; this concentration was selected by performing NC titration during extension) for 3 min at 37°C in 50 mM Tris-HCl (pH=8.0), 80 mM KCl, 6 mM MgCl_2 , 1 mM DTT, and

1) <u>PPT</u> :	5'-AAAAGAAAAGGGGG-3'	$T_m = 46.93^\circ\text{C}$
	3'-CATCTAGAATCGGTGAAAAA TTTTCTTTTCCCCCT GTACCTTCCCGATTAAGTGA-5'	
2) <u>U15</u> :	5'-AAUUCGAGCUCGGUA-3'	$T_m = 58.02^\circ\text{C}$
	3'-TATGCTGAGTGATATCCCGCT TAAGCTCGAGCCAT GGGCCCTAGGAGATCTCAG-5'	
3) <u>U20</u> :	5'-GGGCGAAUUCGAGCUCGGUA-3'	$T_m = 68.94^\circ\text{C}$
	3'-TATGCTGAGTGATAT CCCGCTTAAGCTCGAGCCAT GGGCCCTAGGAGATCTCAG-5'	
4) <u>GP18</u> :	5'-AGGGAAUUUCUUCAGAG-3'	$T_m = 50.13^\circ\text{C}$
	3'-GGAAGGGTGTTCCTTCCGG TCCCTAAAAGAAGTCTC GTCTGGTCTCGGTTGTC-5'	
5) <u>GP18DNA</u> :	5'-AGGGAATTTTCTTCAGAG-3'	$T_m = 64.67^\circ\text{C}$
	3'-GGAAGGGTGTTCCTTCCGG TCCCTAAAAGAAGTCTC GTCTGGTCTCGGTTGTC-5'	
6) <u>PPT3'G/C</u> :	5'-AAAAGAAAAGCGCGC-3'	$T_m = 48.38^\circ\text{C}$
	3'-CATCTAGAATCGGTGAAAAA TTTTCTTTTTCGCGCGT GTACCTTCCCGATTAAGTGA-5'	

Table 2: **List of primers used with their corresponding templates.** The position of each primer on its template after hybridization is shown in bold. The melting temperatures (T_m) for these hybrids was calculated using the “*MELTING*” software (86).

100 μ M dNTPs in a volume of 10.5 μ l, unless specified otherwise. Reactions were initiated by adding HIV-RT wild type or E478Q (2 μ l, ~200 nM final) and incubations were continued for 1-20 min as indicated. The reaction was stopped with 8 μ l of 25 mM EDTA (pH=8.0) and treated with proteinase K for 1 hour at 56°C. Sample buffer (2X: 90% formamide, 10 mM EDTA, 0.25% each bromophenol blue and xylene cyanol) was then added and the samples were electrophoresed on 12% denaturing polyacrylamide gels (7M urea, 19:1 acrylamide:bisacrylamide) as described (110). Products were detected and quantification was performed using a BioRad model FX phosphoimager. The same approach was used for internal labeling assays with the exception of using α -³²P dATP instead of the 5' label. In those reactions ½ of the material was treated with RNases A (0.25 μ g) and T1 (50 units) by incubating at 37°C for 1 min, then increasing to 100°C at a rate of 3° per min, followed by 37°C for 10 min. For internal labeling experiments, excess nucleotides were removed using a G-25 sephadex spin column before electrophoresis.

NC pre-incubation time course assay: Reactions were carried out essentially as described above with the following changes: (1) Pre-incubation times with NC were varied from 1-20 min as indicated; (2) no dNTPs were included; (3) cleavage was initiated by addition of RT and incubation was continued for 1 min before termination with EDTA.

Filter Binding Assay: Filter binding assays were done to study the extent to which the primer-template hybrid bound to nitrocellulose filter membranes in the

presence of various amounts of wild type or NC mutants. The primer (GP18, see Table 2) was 5' end labeled and hybridized to template (1:1 primer:template) as described above. For each reaction, 6 nM prime-template was incubated with the appropriate dilution of NC (0.0625-2 μ M) in the presence of 50 mM Tris-HCl (pH=8.0), 1 mM DTT, 6 mM MgCl₂, and 80 mM KCl in a total volume of 10 μ l at room temperature for 5 minutes. The entire reaction was spotted onto nitrocellulose filter membrane discs (discs were presoaked for 15 min in buffer containing 50 mM Tris-HCl (pH=8.0), 1 mM DTT, 6 mM MgCl₂, 80 mM KCl, and 25 μ M ZnCl₂). The membranes were washed under vacuum 3 times with 1 ml of wash buffer containing 10 mM Tris-HCl (pH=8.0) and 10 mM KCl. The filters were air dried and counted using an LS-6500 Beckman Coulter Scintillation Counter. The fraction of total substrate bound to the filter was calculated and plotted vs. [NC].

Generation of long RNA template: The 430 long RNA fragment without the PPT insert (-PPT) corresponding to part of the *gag-pol* region of HIV-1, was generated from a PCR DNA product derived from pNL4-3 plasmid using primers flanking the region (1894-2323 of HIV pNL4-3 proviral DNA). The primer sequences are as follows: Forward primer 1: 5'-GATTTAGGTGACACTATAGCAAGTAACAAATCCAGCTAC-3'; Reverse primer 1: 5'-AATAGAGCTTCCTTTAATTG-3'. The underlined nucleotides correspond to the SP6 promoter region. PCR reactions included 25 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min, followed by 72°C for 5 min. The products were electrophoresed on 6% native polyacrylamide gels prepared as described (110),

located by UV-shadowing and excised then eluted in 10 mM Tris-HCl (pH=8.0), 1 mM EDTA (pH=8.0). Material was recovered by ethanol precipitation.

Approximately 1/3rd of the material was used in an in vitro transcription reaction using SP6 RNA polymerase from Roche Diagnostics according to the manufacturer's protocol. This material was quantified using UV spectroscopy.

In the +PPT template, the PPT and 5 flanking bases on each side (total of 25 bases) were used to replace 25 nts (bases 2093-2117) in the middle of the above region of pNL4-3 by carrying out two sets of PCRs such that the PPT and flanking regions would be overlapping in each set. The following primers were used to create these fragments in addition to the primers listed above. The PPT and flanking region are italicized: Forward primer 2 (5' – *TTTTTAAAAGAAAAGGGGGGACTGGAAGGCCAGGG*-3') was used with reverse primer 1 (listed above); Reverse primer 2 (5'- *CCAGTCCCCCTTTTCTTTTAAAAACGTAAAAAATTAGCCTGAC* - 3') was used with forward primer 1 (listed above). PCR products were gel purified and the two fragments were joined using the Phusion HF kit from New England Biolabs and were cloned into pBSΔPvuII vector cut with *Hinc* II. GC5 competent cells were then transformed using this vector and grown on LB agar plates with 50 µg/ml carbenicillin. Colonies were picked and plasmid extracted using a Qiagen Miniprep Kit. A colony with the correct sequence was expanded using a Qiagen Midiprep Kit. RNA was produced from PCR DNA as described above for the -PPT template using forward primer 1 and reverse primer 1.

Extension over the long RNA fragment and analysis of 2nd strand priming:

Reactions were carried out using primer-template (10 nM final of 430 nt template) that was pre-hybridized (3:1 primer:template) as described above. Reactions (125 µl final volume) contained 50 mM Tris-HCl (pH=8.0), 80 mM KCl, 1 mM DTT, 6 mM MgCl₂, and 100 µM dCTP, dGTP, and dTTP, and 10 µM radiolabeled α-³²P dATP (0.02 Ci/mmol) in the presence and absence of NC (2 µM). Reactions were initiated by adding wild type HIV-RT (final concentration 200 nM) and continued for 1 hr at 37°C. The RNA fragments generated during DNA synthesis by RNase H cleavage of the template could potentially be used to make a new 2nd strand using the newly made DNA strand as template, (mimicking the generation of the plus strand or 2nd strand during viral replication). Reactions were terminated and digested with proteinase K as described above. This was followed by phenol-chloroform extraction and ethanol precipitation. The samples were then run a short distance on a 6% denaturing PAGE gel and products ~50-320 nts were excised and recovered by overnight elution as described above. This was done to remove the full-length DNA strands made using the RNA as a template, as this product could interfere with subsequent steps. The recovered products were hybridized with a biotinylated primer complementary to the 3' end of the original RNA template (5'-Biotin-AATAGAGCTTCCTTTAATTGCCCCCTATCTTTATTGTGACGAGGGGTCGCTGCCAAAGAGTGATCTGAGGGAAG-3'). Hybridizations used 5 pmoles of the biotinylated primer and an equal amount of total radioactive counts from the – and + NC samples in 55 µl of 50 mM Tris-HCl (pH=8.0), 1 mM DTT, 80 mM KCl as described above. The hybrid was then extended at 37°C for 10 minutes with 1 unit of

Klenow by adding dNTPs (100 μ M) and $MgCl_2$ (6 mM final). This step was performed to remove any 1st strand DNA that remained hybridized to the 2nd strand products using Klenow's strand-displacement activity. The samples were then mixed with 35 μ l of streptavidin agarose that had been pre-treated with the following steps (80): (i) 10 x 1 ml washes using Buffer A (1 M NaCl, 20 mM Tris-HCl (pH=7.3), 5 mM EDTA (pH=8.0), 0.1% NP-40); (ii) Incubate with shaking for 30 min at room temperature with 2 μ g tRNA. After sample addition, material was incubated at room temperature for 30 min with shaking and the following washes were performed: (i) 3 x 1 ml washes with Buffer A; (ii) 6 x 0.7 ml washes with Buffer B (20 mM Tris-HCl (pH=7.3), 0.5 mM EDTA (pH=8.0), 0.1% NP-40); (iii) 6 x 0.7 ml washes with Buffer C (4 M Urea, 10 mM Tris-HCl (pH=7.3), 1 mM EDTA (pH=8.0), 0.1% NP-40); (iv) 6 x 0.7 ml washes with Buffer D (20 mM Tris-HCl (pH=7.3), 0.5 mM EDTA (pH=8.0)). Bound material was eluted in 80 μ l of 2.5 mM biotin solution (2.5 mM Biotin, 1 mM EDTA (pH=8.0)) after incubation at 94°C for 4 min. Half of each sample was treated with RNases A (0.25 μ g) and T1 (50 units) by incubating at 37°C for 1 min, then increasing to 100°C at a rate of 3° per min, followed by 37°C for 10 min. Sample buffer (2X) was added and material was run on a 6% denaturing PAGE gel until the xylene cyanol approached the bottom of the gel. A second aliquot of the same sample was then loaded and electrophoresis was continued until the bromophenol blue dye approached the bottom of the gel. This double loading was performed so that the band shifting due to RNase treatment could be observed for both small and large products. Dried gels were visualized using a Kodak phosphor imaging screen and a BioRad FX imager.

2.3 Results

2.3.1 NC inhibits priming with non-PPT RNA primers by HIV-1 RT

To study the effect of NC on priming by various RNA fragments, we used the primers and their corresponding templates as listed in Table 2. The primers were hybridized to 55 nt long templates such that both ends would be recessed and the product of complete extension would be 35 nts long for all the primers except U20 which would produce a 40 nt product (see Table 2). For HIV derived primers (PPT, GP18, PPT3'G/C), the 55 nt templates were sequences from the corresponding regions of the HIV genome (pNL4-3). The other primers (U15 and U20) were based on sequences from plasmid pBSM13+ (Stratagene). The melting temperature (T_m) for the hybrids formed between each primer and its template are listed in Table 2 and were determined using the "MELTING" program (86). All the primers had higher melting temperatures than the PPT, although the PPT3'G/C was comparable.

Extension of the PPT was compared to several RNA primers in Fig. 7. For the reactions, 5' end labeled primers were hybridized and allowed to undergo extension with RT in the presence or absence of NC (2 μ M final) for increasing amounts of time. For all primers, extension with Klenow polymerase (which extends both RNA and DNA primers), and RNase H cleavage (using HIV-RT in the absence of dNTPs) were used to assess the proportion of primers bound to the template. Reactions were stopped with 10 mM EDTA and the products were run on a 12% denaturing gel after proteinase K treatment (see Methods). Any cleavage or extension products that retain the 5' end of the RNA can be detected by this approach. The PPT was extended but not cleaved in the reaction that included RT and dNTPs (Fig. 7A). Levels of full-

length extension products remained essentially constant from 1 to 20 min. Approximately 70-75% of the total extendable PPT primer (as determined by Klenow extension) was extended in the presence or absence of NC. Some cleavage was observed during prolonged incubation with RT in the absence of dNTPs (Fig. 7A, lane 3). Inclusion of NC did not significantly affect the level of extension over time. For the PPT primer, a portion of the primer typically ran above the fully extended material, consistent with G-quartet formation for this primer (see Fig. 7A left panel) (54). This was unique to the PPT primer. The level of G-quartets was unaffected by NC and did not vary with time. Quartet formation could result in a lower proportion of templates being primed in PPT reactions. However, it is unlikely to influence how NC affects the reaction and these results showing that NC does not influence PPT extension, are consistent with those presented below (see Fig. 8).

One of the other RNAs tested was an 18-mer derived from the frameshift region of the *gag-pol* sequence of the HIV-1 genome (Fig. 7D, GP18). This primer was chosen because it represents an RNA that could potentially be produced during stalling at a strong pause site in the frameshift region (36). As expected, GP18 was mostly cleaved by RT, but some extension was observed (Fig. 7D, 20 min time point, 17% and 2% extension products and 83% and 98% cleavage products for – and + NC, respectively). In the presence of NC, the extension of this primer was strongly inhibited by approximately 90% at the longest time point. The selective inhibition of GP18 relative to the PPT was not length dependent as extension of a random unrelated 15-mer (U15, Fig. 7C) and a 20-mer which was identical to the 15-mer at the 3' end but included 5 additional 5' nts (U20, Fig. 7E), were also inhibited by NC

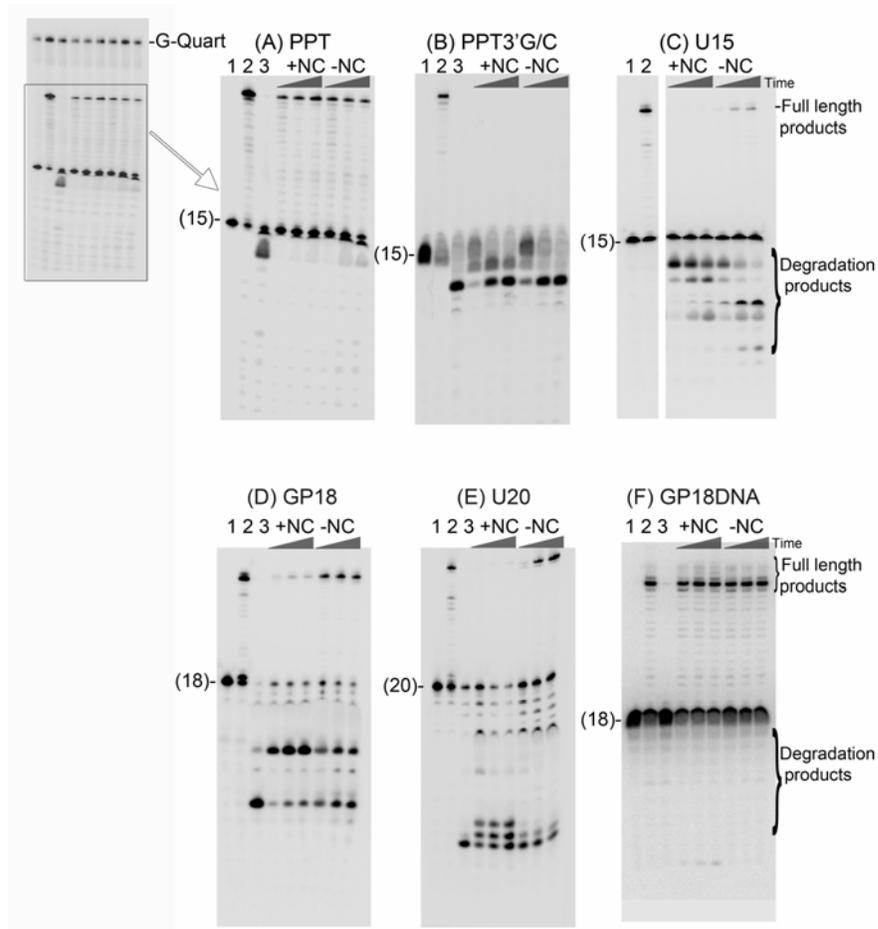


Figure 7: **Nucleocapsid protein inhibits non-PPT RNA primer extension by HIV-RT.** Six different primers (Panels A–F and see Table 2) were extended by RT in the presence and absence of NC (as indicated) for increasing amounts of time (1, 10 and 20 min). The figure shows the images developed after running the samples on a 12% denaturing PAGE gel. For the PPT reactions (panel A), an inset gel showing the G-quartets (G-Quart) that migrate above the fully extended products are shown (see Results). All reactions were repeated 2 or more times. Migration positions (marked by primer size in nts) of full-length primers are indicated in each panel as are positions of fully extended and degradation products. Lane 1, primer control with no RT or NC; lane 2, full extension control (incubation with RT E478Q for 20 min in the absence of NC); lane 3, RNase H control (20 min incubation with HIV-RT in the absence of NC and dNTPs, see Fig. 10C for an RNase H control for U15).

(>95% inhibition of extension for both U15 and U20). The patterns were similar to those observed with GP18, especially for U20. For U15, the level of extension products was lower with less than 3% extended (of total cleavage and extension products) even in the absence of NC. Since the predicted T_m s of the non-PPT primers were considerably higher than the PPT (Table 2), an additional control using a primer with a similar T_m was performed. Primer PPT3'G/C was similar to the PPT except 3 of the 6 G residues were replaced by Cs at the 3' end, thus maintaining the G/C content. This primer was cleaved by RT but no extension products were observed even in the absence of NC, although RNase H minus HIV-RT (E478Q) did extend it (see below). In contrast to these results, a DNA version of the GP18 primer was efficiently extended by RT and unaffected by NC (Fig. 7F). These results show clearly that NC inhibits the extension of non-PPT RNA primers by RT.

Nucleocapsid protein is known to induce aggregation of nucleic acids under conditions similar to those employed in these experiments (8). Since this could potentially influence the results, NC's ability to aggregate the primer-templates (labeled template strand) was tested by centrifugation after preincubation with NC as described (8). A low level of aggregation (~20% of substrate pelleted when NC was added) was detected but the vast majority of the substrate (80-90%) remained in the supernatant fraction (data not shown). The low aggregate formation in these experiments may be due to the nature and amount of nucleic acid used here in comparison to the previous work. In this work 6 nM of a 55 nt DNA template was used while ~16 nM of RNA templates that were ~ 1500-4000 nts were used in the previous work.

2.3.2 Internal labeling experiments indicate that all non-PPT RNA priming events are inhibited by NC:

Since the primers in the above experiments were 5' end labeled, only those products which retain the 5' end of the RNA were detected. It is possible that some products were extended from RNA cleavage fragments derived from the 3' end of the primer or the labeled RNA was cleaved after RT DNA extension. Whether RT extension of these is inhibited by NC cannot be assessed using 5' end labeling. To determine if NC affected RT extension of these products, the next set of experiments were carried out using α -³²P dATP in order to internally label the products (see Methods). Three RNA primers (PPT, U15 and U20) were extended for 20 minutes with RT on their respective templates in the reactions. Half of each sample was treated with RNase A and T1 to remove any RNA from the primer still attached to the product. The length of the DNA portion of the products in turn would help us detect where priming actually began after cleavage. A Klenow control was used in each case to detect full extension of the uncleaved RNA. Klenow can add dNTPs to the 3' RNA terminus to produce fully extended products. Complete cleavage by the RNases would then remove the RNA leaving only the DNA nts (and perhaps 1-2 RNA nts) that were added to the primer. For all primers, 20 nts should be added to produce full-length products if extension occurred directly from the 3' end of the intact primer. For PPT extension with RT, the full-length products observed in Fig. 7A were also seen, but in addition, some smaller products, a few nucleotides longer than the primer, were also observed (Fig. 8A). These were probably DNAs produced by RT extension followed by excision of the PPT primer. No change in the PPT reactions was

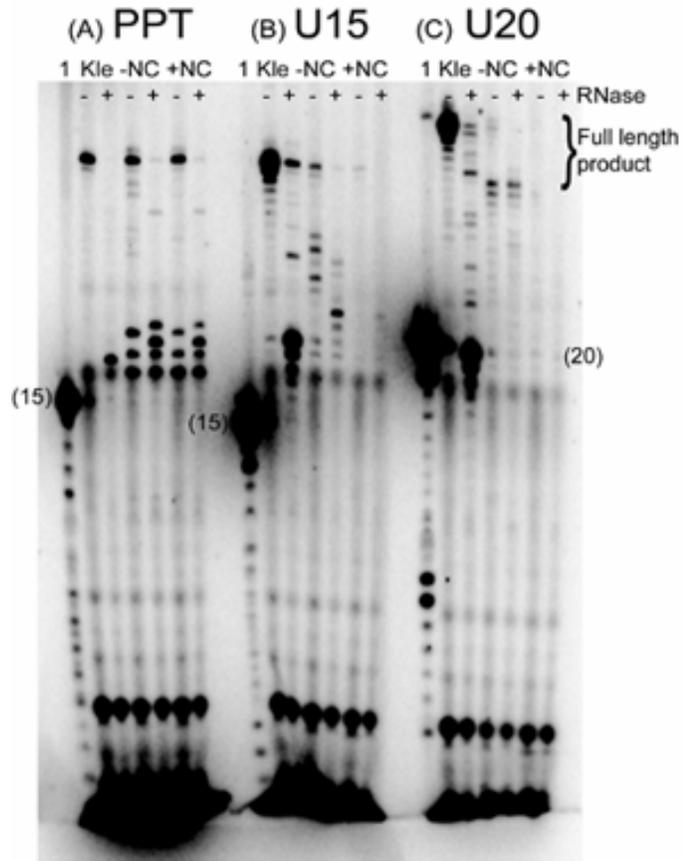


Figure 8: **Autoradiogram of primer extension by RT using internal labeling.** Three primer-templates, PPT, U15 and U20 were used as indicated (see Table 2). Lane 1, primer control using 5' labeled primer. Klenow (Kle) reactions were in the absence of NC. RT extensions were – or + NC as indicated. Half of each sample was treated with RNases A and T1 (+) while the other half was not (–). Positions of intact primers are indicated by the primer size in nts and the position of fully extended products is indicated. The experiment was repeated twice to confirm results.

observed with NC. With U15 in the absence of NC, several extension products were observed, most of which were digested by RNases down to a single major product of ~22 nts and other slightly longer and shorter products. The 22 nt product was 2 nts longer than the Klenow digest. This suggests that many of the DNAs produced from U15 were derived from a 13 nt RNA that had 2 nts removed from the 3' end prior to extension. The smaller extension products observed without RNase digestion probably resulted from cleavage at different positions after extension by RT. Unlike the PPT reactions, there was a strong inhibition of all extension in reactions with NC. With U20, the major extension product observed without NC was ~31 nts along with a smaller amount of fully extended product. Only the fully extended material was observed in the 5' end labeling experiments (Fig. 7E). The full length, but not the 31-mer disappeared after RNase treatment. Our interpretation of this result is that RT first cleaved U20 and then extended the 5' derived cleavage product which was ~9 nts long. Some of the products then had the 9 nts of RNA removed by RT RNase H activity. However, the possibility that these are intermediate extension products that were not evident in the 5' end label experiments because they did not retain the 5' label, cannot be ruled out. Like U15, NC strongly inhibited all extension in the reactions. These results indicate that the inhibition observed with NC is universal in the sense that all priming events occurring during extension of non-PPT primers are inhibited by NC.

The most obvious reason for such an inhibition pattern would be the helix destabilizing activity of NC (see Introduction) causing dissociation of short hybrids. An alternative explanation is that NC, a nucleic acid binding protein, directly blocks

the binding of RT to the RNA 3' terminus thus impeding extension. This is especially possible since RT binds poorly to RNA primers in the orientation that allows extension (39, 42, 43, 97). To study which of these properties of NC plays a role in the inhibition of non-PPT primers, the experiments below were performed.

2.3.3 NC does not dissociate intact RNA primers from the template but can destabilize smaller fragments attached to the template after RT cleavage:

To see if the helix destabilizing activity of NC plays a role in the selective extension of PPT, a time course assay with NC and GP18 was performed. Here, the primer-template hybrid was incubated for increasing amounts of time with NC before adding RT in the absence of dNTPs to detect cleavage. A decrease in cleavage would show that the primer was no longer associated with the template indicating that helix destabilization activity of NC caused dissociation of the primer from the template. In all reactions when RT was added, a cleavage product at ~10 nts was present and no decrease in the level of cleavage of the 18-mer over the course of the incubation periods (1 to 20 minutes) was observed (Fig. 9). However, after cleavage with RT, in the -NC control reaction two fragments, the smaller of which (~ 5 nts) was only faintly visible in the lanes with NC, were observed. We also tested U15 and obtained similar results (data not shown). This would suggest that once the primer was cut, NC caused its dissociation from the template preventing further degradation by RNase H while in the absence of NC some cleavage products remained bound to the template and were cleaved a second time. This has also been observed previously (125). In addition, this pattern is also observed in Fig. 7 with primers GP18 and U15, where

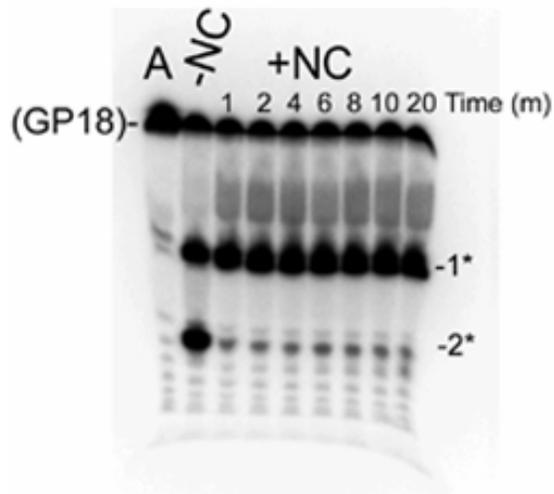


Figure 9: **NC causes dissociation of cleavage products but not uncleaved primers.** 5' end labeled primer GP18 bound to 55 nt template (see Table 2) was incubated with NC at 37 °C for increasing amounts of time (1, 2, 4, 6, 8, 10, and 20 min as indicated). HIV-RT was then added and incubation was continued for 1 min. Lane A is primer control without RT or NC. The – NC lane is a cleavage reaction with RT but without NC for 1 min. The position of uncut primer (GP18) and the major cleavage products (1* and 2* are the primary and secondary cleavage products, respectively) are designated. The assay was repeated to confirm results.

smaller, presumably secondary cleavage products are inhibited in the presence of NC. Taken together, these results imply that NC does not cause dissociation of longer primers, but once the primers are cut by RT, the resulting smaller fragments may be dissociated by NC's unwinding activity and this could be one mechanism for inhibiting priming by non-PPT RNA oligonucleotides that are more prone to cleavage by RT than PPT.

2.3.4 NC inhibits priming by directly preventing extension from the 3' RNA terminus:

To study if NC could cause inhibition of non-PPT primers in the absence of cleavage and subsequent dissociation, extension with an RNase H deficient mutant of RT (E478Q) was performed. This mutant with a glutamate to glutamine substitution at position 478 has earlier been found to have polymerase activity similar to wild type (112). If helix destabilization was the only mode of inhibition by NC, extension levels with E478Q in presence and absence of NC would remain the same, since the primer wouldn't be cut into smaller fragments that could then be dissociated by NC. Reaction conditions were identical to those in the extension assay with wild type RT. Results showed that NC potently inhibited RNA extension by E478Q for all the non-PPT RNA primers tested (Fig. 10). As expected, no cleavage was observed. Therefore all extension products could be observed as the 5' label would still be attached. Also, as none of the primers are degraded by the enzyme, more extension was observed compared to wild type (see Fig. 7). Extension of PPT by E478Q was unaffected by

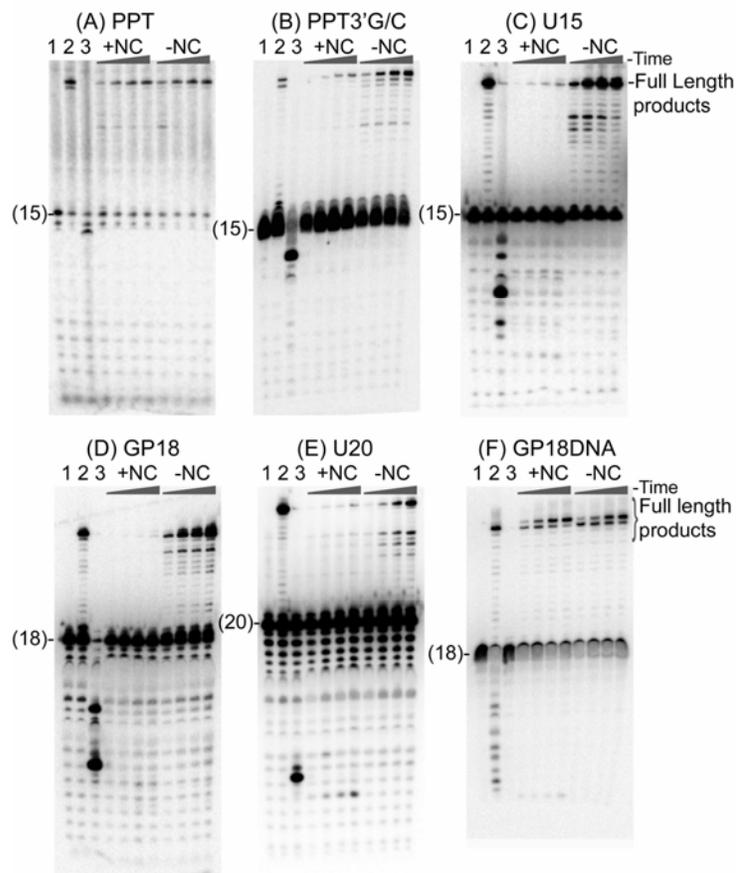


Figure 10: **NC inhibits extension of full-length non-PPT RNAs.** Six different primers (see Table 2) were extended by E478Q (RNase H minus HIV-RT) in the presence and absence of NC. Lengths of the primers are marked in the image. Reactions were carried out in the presence and absence of NC as indicated for increasing amounts of time (1, 5, 10, and 20 min). The figure shows the image developed after running the samples on a 12% denaturing PAGE gel. All reactions were repeated 2 or more times. Lane 1 for each set is a primer control (reaction without E478Q or NC). Lane 2 is a full extension control where the hybrid was extended by 1 U of Klenow for 20 min in the absence of NC. Lane 3 is an RNase H control where reaction was carried out for 20 min in the absence of NC and dNTPs.

the presence of NC just as it was with wild type RT, while ~ 90% inhibition of extension was seen in the case of non-PPT RNA primers. The DNA primer used (GP18DNA) remained relatively unaffected by the presence of NC (Fig. 10F). These results suggest that NC is able to inhibit RNA primers from being extended even when they remain hybridized to the template. RT's preferred binding mode to RNA primers (other than PPT) strongly favors cleavage and the polymerase domain orients near the 5' end of the primer (39, 42, 43, 97). Binding in a mode that favors extension (i.e. polymerase domain near the RNA 3' end) is relatively weak. Therefore it is conceivable that NC would compete with RT for binding to the 3' ends and thus block extension. Since the PPT is presumably designed to bind RT with its polymerase domain at the 3' end, it would be more resistant to inhibition by NC. To further elucidate the role of competitive binding by NC, experiments using a mutant of NC that was deficient in binding were performed.

2.3.5 An NC mutant that shows weaker binding to nucleic acid shows reduced inhibition of RT RNA priming:

Extension assays in the presence of NC mutants with varying binding capacities were performed. Three different types of NC proteins were used (two mutated proteins, 1.1 and 2.2 along with wild type NC). In 1.1 NC there are two copies of zinc finger 1 with finger 2 being replaced by finger 1, while zinc finger 1 is replaced by 2 in 2.2 NC (60). Binding of each NC was tested on the GP18 primer-template hybrid using nitrocellulose filter binding assays. Wild type NC bound slightly better than 1.1 while 2.2 bound the substrate much more weakly (Fig. 11D, a

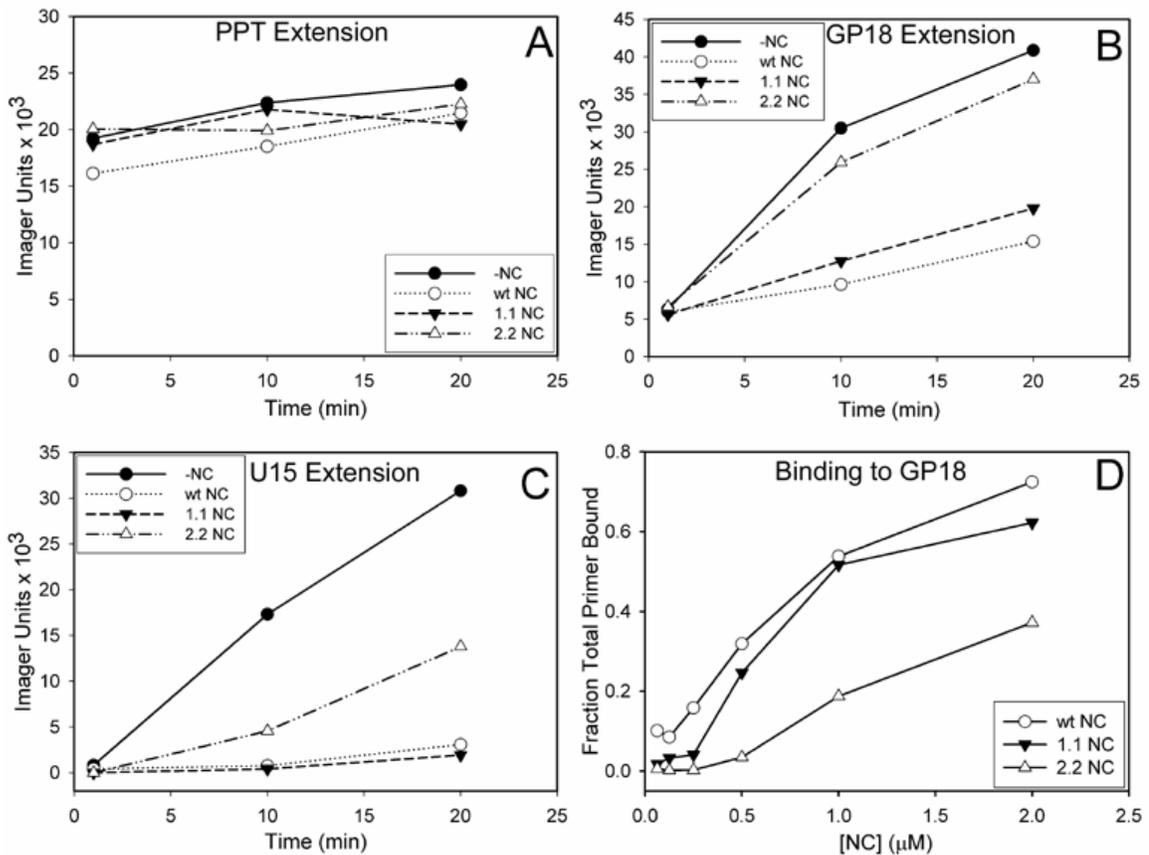


Figure 11: **Graphs of RNA primer extension in the presence of NC mutants** show that mutants that bind nucleic acid less tightly show less inhibition of RNA primer extension. Panels A, B, and C show extension of PPT, U15 and GP18 primer-templates (see Table 2), respectively. Extension was carried out in the presence of no NC, wild type (wt), 1.1, or 2.2 NCs over 1–20 min. Plots show imager units (counts \times mm²) vs. time. (D) Plot of binding of the different NC mutants 1.1 (filled triangle), 2.2 (open triangles) and wild type (open circles) with the GP18 primer-template. The plot shows the relative amount of total material in the reaction that bound to a nitrocellulose disk vs. increasing concentrations of NC. All experiments were repeated at least once and representative graphs are shown.

similar results was obtained with U15 primer template (data not shown)). In extension assays, elongation of the PPT remained fairly constant in presence of all three NCs (Fig. 11A), while inhibition of GP18 (Fig. 11C) extension by 1.1 was comparable to wt and 2.2 was showed less inhibition. For primer U15 (Fig. 11B), similar results were observed except that 2.2 was even less effective as an inhibitor. The results indicate that NC proteins that bind with higher affinity to the substrate are better inhibitors of RNA extension, consistent with the hypothesis that NC competes with RT for binding to the primer 3' terminus.

2.3.6 NC inhibits non-PPT RNA priming on a long RNA template:

To study whether extendable primers are generated by RT during reverse transcription, a 430 nt long RNA fragment derived from the *gag-pol* frameshift region of the HIV-1 genome (1894-2323 of HIV pNL4-3 proviral insert) was used as a template (-PPT template in Fig. 12). Extension over this template can potentially give rise to RNA fragments that could be used for second strand priming. As a control, the central region of this template was substituted with the 3' PPT and the five nucleotides flanking each side (+PPT template in Fig. 12), to see if incorporating this sequence would focus second strand priming to this region. A 20 nt DNA was used to prime synthesis over these templates. Primer extension with wild type HIV-RT was performed in the presence of internal label. Fragments generated by the RNase H function of RT could be used for priming over the newly made DNA strand. To detect these “2nd strand” products a 75 nt oligonucleotide that was biotinylated at the 5' end

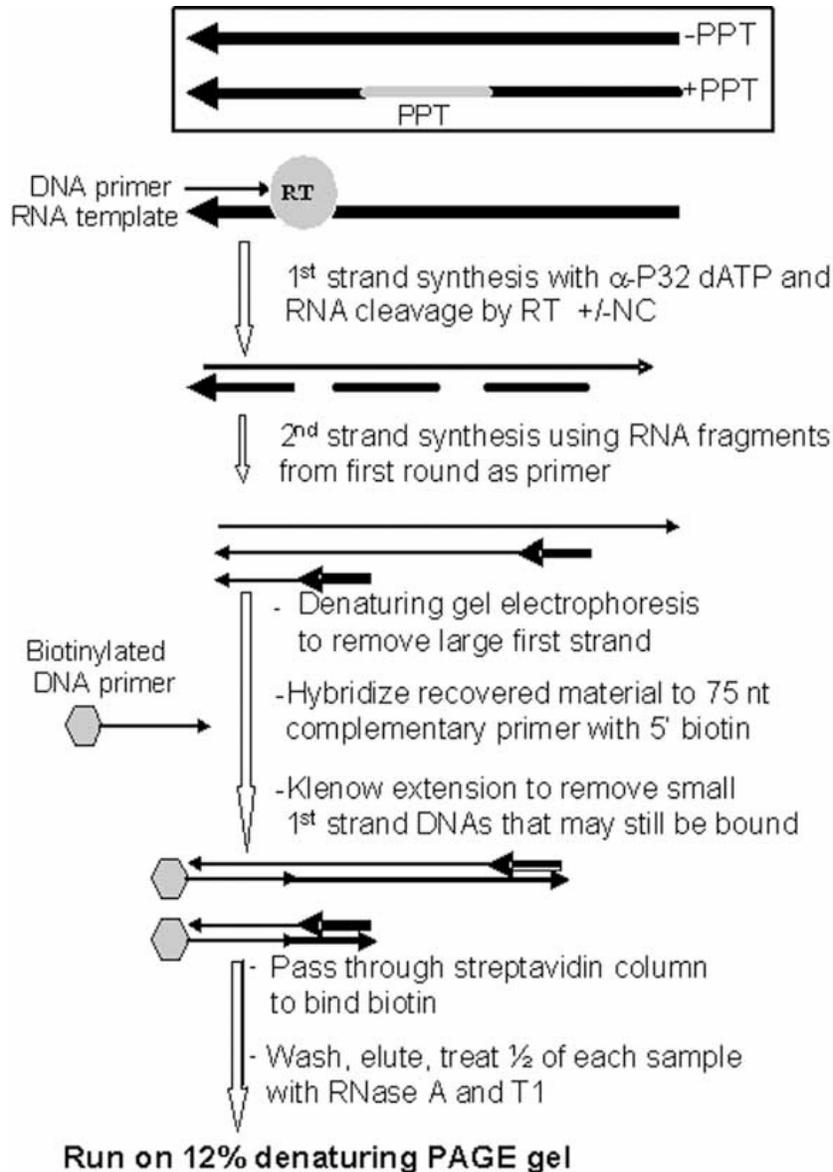


Figure 12: **Schematic representation of protocol for detecting RNA primed 2nd strand synthesis on a long RNA template in the presence and absence of NC.** Refer to Materials and Methods for details. Top box: representation of the two 430 nt templates used in the experiment (+ and – PPT) with 25 nt PPT region shown in gray. Nucleotide positions of the template strand are indicated.

was used. This oligonucleotide would be complementary to any DNA products originating from RNA fragments that were subsequently extended to the end of the first strand DNA (see Fig. 12). Material from the reactions was run on a denaturing gel and products shorter than approximately 300 nts were excised and eluted, then hybridized to excess biotinylated oligonucleotide. This was followed by extension with Klenow. This served two purposes: (1) extension of the biotinylated oligonucleotide using the 2nd strand product as template would result in a longer more stable hybrid; (2) any first strand DNAs that remain associated with the 2nd strand products would be dislocated by the strand displacement activity of Klenow and therefore not interfere with the final analysis. The biotinylated hybrid was then passed through a streptavidin column to bind the biotin-associated products and washed several times to eliminate non-specific binding. Products were run on a 6% denaturing gel after half of each sample was treated with RNases A and T1. Results showed products with both the – and + PPT templates in absence of NC (Fig. 13). Products in -PPT reactions were more diffuse and greatly suppressed in the presence of NC. RNase treatment caused a downward shift in most bands in the –PPT reactions indicating that they had retained part of the RNA primer from which they originated. With the +PPT template, the profile was dominated by a distinct band that migrated consistent with priming from the location of the PPT. The intensity of this band did not change even in presence of NC. The band also did not shift downward after RNase treatment indicating that the PPT was removed subsequent to its use as a primer. Some less intense bands that ran below the PPT product were also observed in +PPT reactions. Some of these bands showed small shifts after RNase treatment

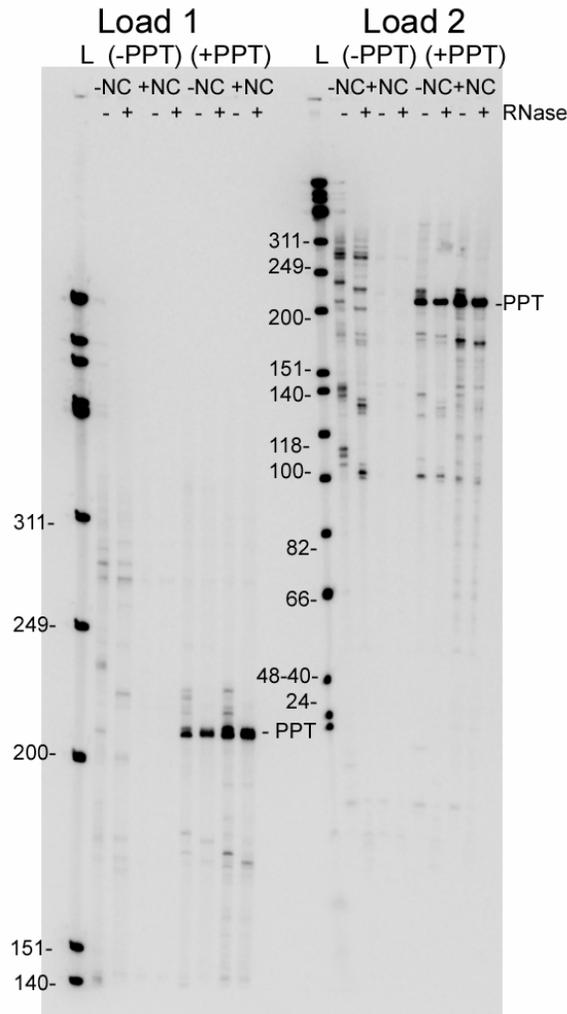


Figure 13: **NC inhibits priming by non-PPT RNA primers generated by RT during extension on a long RNA template.** Material isolated from the protocol described in Fig. 12 was run on a 6% denaturing polyacrylamide gel as described in Materials and Methods. In order to visualize both large and small products and changes in product migration after RNase treatment, samples were loaded twice with Load 1 loaded first followed by electrophoresis and reloading of samples (Load 2). The + and – PPT templates are indicated. Each template was extended in the presence and absence of NC as indicated. Half of each sample was treated with RNases A and T1 as indicated. ϕ X174 ladder (L) digested with Hinf I was used as the molecular size marker with sizes of select bands indicated. The position of 2nd strand DNAs initiated from the PPT primer (PPT) is indicated. The assay was repeated twice to confirm results.

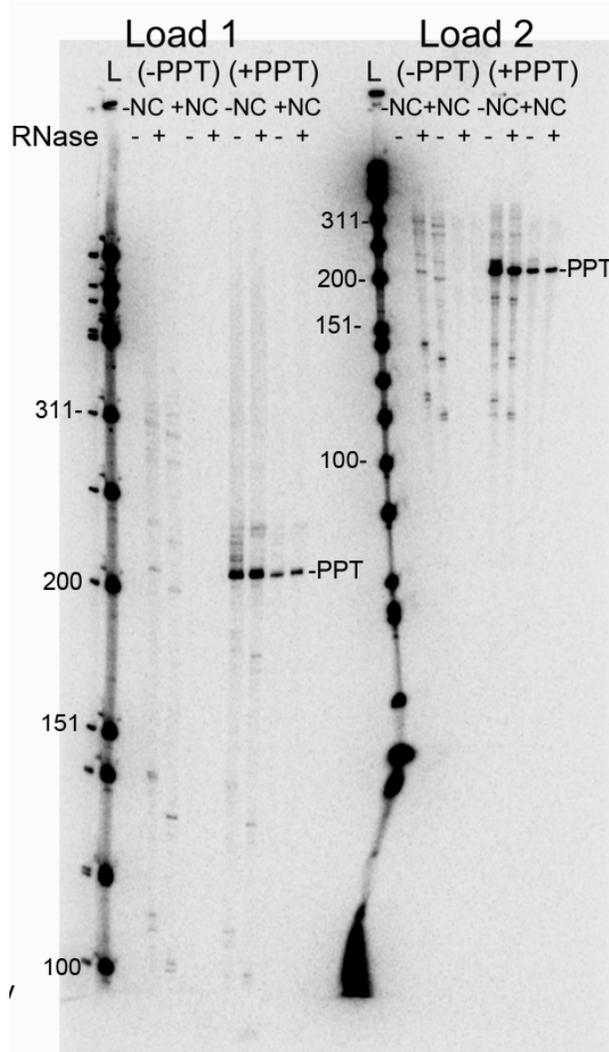


Figure 14: **NC inhibits non-PPT RNA primer extension over a long RNA template at 1 mM Mg²⁺**. The figure is performed as shown in Fig. 12 using 1 mM MgCl₂ rather than 6 mM as in the previous figure. This figure shows that for the -PPT templates, 2nd strand priming also occurs at 1 mM Mg²⁺ and is strongly inhibited by NC. As in Fig 13, samples were loaded twice with Load 1 loaded first followed by electrophoresis and reloading of samples (Load 2). The + and – PPT templates are indicated. Each template was extended in the presence and absence of NC as indicated. Half of each sample was treated with RNases A and T1 as indicated. ΦX174 ladder (L) digested with Hinf I was used as the molecular size marker with sizes of select bands indicated. The position of 2nd strand DNAs initiated from the PPT primer (*PPT) is indicated. The assay was repeated twice to confirm results.

(see +PPT Load 1) although the extent of shifting was less than observed with 2nd strand products from the –PPT template. Curiously, these products were not inhibited by NC. This experiment was also conducted using 1 mM MgCl₂ (Fig. 14), a conditions that would lead to tighter NC binding (127). The level of 2nd strand priming on the –PPT template was lower and strongly inhibited by NC. Priming from the PPT on the +PPT template was also modestly inhibited by NC. In contrast to reaction at 6 mM MgCl₂, smaller products on the +PPT template were also inhibited in the presence of NC suggesting that tighter NC binding is required to eliminate these 2nd strand priming events. Because these products behave similar to PPT-derived products, one possibility is that they were also generated from the PPT but were truncated or incompletely extended.

2.4 Discussion

In this study I show for the first time that the chaperone protein, HIV NC, may play a role in the choice of RNA primers for 2nd strand synthesis by HIV-RT. The effect of NC on priming was examined using PPT and non-PPT primers. Confirming earlier studies, RT was capable of non-PPT RNA primer extension, albeit, poorly (see chapter introduction). As expected, these primers were rapidly cleaved in reactions with wild type RT (Fig. 7) and internal labeling experiments indicated that a small portion of the intact primer (most evident for U15 (Fig. 8B)) or cleavage products (most evident with U20 (Fig. 8C)) were extended by RT. However, in the presence of NC, non-PPT priming was greatly reduced while PPT priming was unaffected (Fig. 8). NC also had an effect on the cleavage profile but no obvious effect on the level of total primer that was cleaved (see Results and Fig. 9). This inhibition was also evident in a more dynamic reaction that tested priming over a long RNA template (Fig. 13). The mechanisms used to inhibit priming included increasing the rate of dissociation of RNA cleavage products (see Results) and directly inhibiting extension from the 3' terminus, most likely by blocking binding.

It was clear from experiments using RNase H minus HIV-RT (E478Q), that NC directly inhibited extension from the primer 3' end (Fig. 10). As stated in the Results, the most likely explanation of this is direct competition for binding between RT and NC. The fact that 2.2 NC, which binds with lower affinity than wild type or 1.1, was less inhibitory supports this hypothesis. However, other differences between these NCs cannot be excluded as potentially contributing to the results. NC 2.2 has low helix-destabilizing activity compared to wild type and 1.1 (64, 66). Since

experiments with these NCs were carried out with E478Q, dissociation of cleavage products, which would likely be more prominent with wild type and 1.1, would not have been a factor as NC was found to be incapable of dissociating full-length primers (Fig. 9). Despite this, we can not rule out the possibility that helix destabilizing activity rather than binding affinity for nucleic acids played a role in the results, although it seems unlikely. Another possibility is that NC inhibition is related to direct binding interactions between RT and NC which have been reported (121). It is possible that the 3 different NCs used associate differently with RT. Nevertheless, competitive binding between RT and NC is the most reasonable explanation of the data. The lack of potent inhibition by NC with a DNA primer could then be explained by RT's higher affinity for the DNA 3' end in contrast to low affinity for RNA 3' recessed termini (see Results). We also tested other NC point mutants known to bind with different affinities to nucleic acid including I24Q, N27D, N17K, and F16W (94). In general, the level of RNA primer inhibition with these mutants was also consistent with binding affinity for the substrate (data not shown). Although all mutants that bound more tightly inhibited RNA extension more prominently, these results cannot unequivocally differentiate between direct binding inhibition and inhibition related to helix-destabilizing activity. All mutants with lower binding also show decreased helix destabilizing activity and it is unlikely that any mutation that lowers binding would not affect destabilizing activity. Our efforts to prove direct binding inhibition more directly by measuring K_d values of RT for RNA primers in the presence and absence of NC were unsuccessful. Assays with the RNA primed templates performed using an RT "trap" to sequester the enzyme (E478Q) after a single round of binding and

extension produced no extension products even with high enzyme concentrations. This may be due to an extremely high K_d such that it was not feasible to add enough enzyme, combined with the trap affecting extension of pre-bound enzymes.

To create a system that would mimic viral replication more closely than simple short primer-template based experiments, long RNA templates were used. Results obtained support the idea that RNAs generated during minus strand priming can be used for priming plus strand synthesis (Fig. 13). This is in agreement with studies that show that some retroviruses frequently use primers in addition to their PPT for plus strand priming resulting in discontinuous double stranded DNA prior to integration (see Introduction). Even HIV does this to some extent as has been shown in both in cell culture (92) and cell free (78) systems. In our system, several different RNA priming positions were observed over the 430 nt –PPT template indicating that many RNAs can potentially be used to prime 2nd strand synthesis. It is notable that this represents just a fraction of the total length of the HIV genome suggesting that in the absence of NC, RNA priming would occur at numerous positions. RNA priming occurred from discrete locations on the template indicating that some sites were preferred over others. The general effect of NC was to strongly decrease, but not completely inhibit the level of priming from the specific locations. Possible reasons for preferred sites include: (1) a PPT-like character that inhibits cleavage and/or attracts RT to the 3' end; (2) a high G/C content that allows the cleavage product to associate more strongly with the nascent DNA; and, (3) RT pausing in the vicinity to facilitate cleavages that can generate such primers. We plan to determine the sequence of the fragments used for priming on this and other templates to help

understand why they were used. Although both helix-destabilizing and direct inhibition of extension were implicated as factors in decreasing RNA priming, the extent to which each of these contributed to the results shown in Fig. 13 cannot be determined.

As noted above and in the Introduction, some retroviruses use multiple priming sites for plus strand synthesis giving rise to discontinuities in the plus strand that must be resolved before proviral transcription. The extent to which a particular retrovirus uses non-PPT primers for plus strand synthesis could depend on several factors including: (a) the ability of the cognate RT to extend these RNAs; (b) the size of fragments generated by RNase H cleavage during first strand synthesis (in this regard it is interesting that avian myeloblastosis virus RT generates significantly larger fragments than HIV-RT during a single round of RT synthesis (44) and avian retroviruses are also known to produce discontinuous plus strands); and, (c) differences in the level of inhibition of non-PPT RNA priming by the different NC proteins. With regard to (c) and contrast to our findings, others have shown that moloney murine leukemia virus (MuLV) NC protein does not appear to enhance PPT utilization or suppress extension from non-PPT primers (114). This could be due to different chaperone properties on HIV and MuLV NC (49, 124).

Whether the PPT is critical for priming of plus strand synthesis in HIV is unclear although it is clearly required for efficient replication. Others have reported that complete randomization of the HIV-1 PPT still allows for minimal replication of the virus indicating that this primer is not absolutely essential for the virus to survive and that the virus is capable of initiating the plus strand from regions other than the

PPT (91). The randomized PPT, however, reverted back to its original state within a few generations emphasizing the preference of the virus for this sequence.

Interestingly, a major defect of HIV viruses with randomized PPT regions is at the level of integration, highlighting the importance of the PPT for this process (91).

Thus it is difficult to say conclusively if PPT preference results from its necessity as a primer, its role in integration, or in defining the start of the 5' LTR.

The biological significance of HIV NC inhibiting non-PPT RNA priming is yet to be determined. Especially since some other retroviruses do not seem to strongly inhibit this process. The possibility that the inhibition of non-PPT priming by NC has no biological significance but is simply a consequence of NC's properties cannot be dismissed. However, one possible reason for the disparity among retroviruses could be related to the differences in cell tropism, since any plus strand discontinuities must be resolved by the host cell machinery before transcription. Variations in the level of the resolving activities in the different cells could play a role in determining how well discontinuities are tolerated.

Chapter 3: Effect of NC and Mg²⁺ on HIV Recombination

3.1 Introduction

3.1.1 Significance of recombination

A high recombination rate is one of the hallmarks of HIV replication. This as well as an error prone viral polymerase leads to a genetically diverse population, even within patients infected only one time. Hence, HIV is generally referred to as a “quasi-species” indicating the huge diversity within the group. The geographical proximity of certain HIV-1 subtypes and multiple exposures to the virus has made it possible for a single host cell to be infected by two different clades of the virus. Co-packaging of hetero-dimeric RNA in such cases can cause the generation of inter-subtype recombinants. Some of these recombinants survive well in the host and are successful at beginning new and stable infections (CRFs, see section 1.13).

Recombination provides the virus with a powerful tool under conditions detrimental to its growth and spread. For example, the emergence of resistance to intensive treatments such as triple drug therapy requires the acquisition of resistance features to at least three drugs by a single virus. Even with the high error rate of RT the probability of acquiring a set of mutations that lead to triple drug resistance and viability is very low. Therefore such recombinants probably require recombination to be produced in a relatively short time. The hypervariability of HIV also makes it a poor candidate for vaccine development. The lack of an effective cure or competent

immunization strategy and the emergence of resistant mutants to the existing therapeutics, make the study of HIV variants very significant.

3.1.2 Mechanisms of recombination

The process of recombination or strand transfer occurs during the copying of the genomic RNA into double stranded DNA, when the nascent DNA strand switches from one of the viral RNAs (termed “donor” in in vitro reactions) to the other (acceptor in in vitro reactions). For the successful completion of reverse transcription, two end transfers are necessary. The first end transfer occurs when the -sssDNA switches from the 5' end of the genome to the 3' end of the same or different RNA molecule. The second obligatory transfer occurs when the +sssDNA jumps from the 5' end of the minus sense DNA strand to its 3' end. The presence of sequence homology is necessary for strand transfers to occur. Studies by Baird et al. show that in the hypervariable regions of the envelope glycoprotein gp120, recombination between different virus subtypes is greatly reduced and tends to occur in small stretches of homology (10). The majority of strand jumps occur during the synthesis of the first strand of DNA, the minus strand. During this step of the replication reaction, the donor template is degraded by the RNase H activity of RT and the unbound portion of the newly made DNA can easily associate with the acceptor template, a step facilitated by NC protein. Though strand transfers do take place during the synthesis of the plus or the second strand, they occur to a lesser extent (70).

Pausing during DNA synthesis by RT has been shown to have an effect on the probability of recombination in a given region. The polymerase and RNase H functions of HIV RT are uncoupled with polymerization occurring faster than degradation. Hence, when RT pauses, either due to an aberration such as breakage of the template or because it encounters a secondary structure or due to host factors such as low dNTP concentrations, the rate of RNase H hydrolysis increases compared to synthesis and this leads to an increase in degradation of the RNA template. This creates a cleared region where the nascent DNA can associate with the acceptor template to initiate strand transfer. The relationship between the rate of recombination and the balance between the RNase H and polymerization activities of RT forms the basis for the “Dynamic copy-choice model” (72).

Much of what is known about recombination comes from work with reconstituted in vitro assay systems. These systems are easy to manipulate and provide data in a short time frame. HIV proteins required for recombination are also easy to purify in expression systems and show high stability. Some of the findings from these assays have been tested in cell culture recombination systems. In general the results have been consistent. For example, a requirement for RNase H activity and the correlation of some pause sites that are hotspots in in vitro assays with recombination hotspots in cells. Still, it is not possible to completely mimic the environment of the viral core in the cell in a reconstituted assay. Therefore, it is not clear whether all the information garnered from in vitro assays accurately depicts viral recombination.

As noted earlier, NC protein enhances recombination in in vitro reactions. Typically such reactions are performed under conditions that optimize RT activity (~ 80 mM KCl and ~ 6 mM MgCl₂). However, these conditions do not necessarily correlate to cellular conditions particularly with respect to Mg²⁺ concentrations. Although total Mg²⁺ in cells is relatively high (~ 10 mM) and variable depending on cell type, free Mg²⁺, a more important determinant of enzyme activity, is much lower (~ 1 mM, (59)). NC binding has also been shown to be affected by the concentration of Mg²⁺ in the reaction. At low concentrations of the cation, the binding affinity of the protein for nucleic acids is higher. It has been proposed that Mg²⁺ competes directly with NC for binding to nucleic acids and can displace NC molecules. Even in vitro, recent results show that –sssDNA strand transfer is more efficient in the presence of NC when suboptimal (with respect to RT) Mg²⁺ concentrations are used (127). Magnesium also has a pronounced effect on the structure of RNA where it stabilizes secondary structures, leading to increased RT pausing.

A careful study of the interplay between Mg²⁺ and NC in in vitro assays testing internal recombination has not been performed and is the goal of the current work. The ultimate goal is to correlate in vitro findings to results found in cell culture. This could be accomplished by determining under what conditions the locations where recombination occurs in vitro are the same as those observed in cells. Presumably these conditions would most closely mimic the cellular environment. This comparison would not only enable us to ratify our findings, but also help us assess whether recombination can be studied in vitro as reliably as it can be in cells.

3.2 Materials and Methods

3.2.1 Materials

The expression clone for wild type HIV-RT (strain HXB2) was kindly provided by Dr. Samuel Wilson (NIEHS, Research Triangle Park, NC). This enzyme (non-histidine tagged version) was purified as described (67). The expression clone for wild type NC was a gift from Dr Charles McHenry (University of Colorado) and was purified as described (128). Both NC and RT proteins were tested for purity using SDS-PAGE gels (data not shown). Plasmid pNL4-3 was from the NIH AIDS Research and Reference Reagent Program. T4 polynucleotide kinase was from New England Biolabs. Proteinase K was from Shelton Scientific Inc., CT. Plasmid preparation for sequencing was carried out using the Qiagen Miniprep kit. The QuikChange Multi Site-Directed Mutagenesis Kit and the Strataclone PCR Cloning Kit were from Stratagene. Radiolabel was purchased from Amersham and Perkin Elmer. Sephadex G-25 spin columns were from ISC Bioexpress. SP6 RNA polymerase, DNaseI/RNase-free and RNaseI/Dnase-free were from Roche Diagnostics. All oligonucleotides were obtained from Integrated DNA Technologies and all other reagents were from Fisher and VWR.

3.2.2 Methods

Generation of donor and acceptor substrates: Two sets of donor and acceptor substrates were used in this study. The two sets analyzed overlapping regions of the HIV-1 *gag-pol* region. In the first set, the donor was amplified from bases 1924-2343

of pNL4-3 plasmid using forward primer 5'-
GATTTAGGTGACACTATAGTCAAAACAGAAAGGCAATTTTAGGAAC and
reverse primer 5'-TATCATCTGCTCCTGTATCT. The underlined nucleotides
correspond to the SP6 promoter region while the italicized bases are of non-HIV
origin added to prevent the DNA synthesized on the donor from transferring to the
acceptor. The acceptor sequence was also generated from the *gag-pol* region (bases
1894 to 2323) using forward primer 5'-

GATTTAGGTGACACTATAGCAAGTAACAAATCCAGCTAC

(SP6 promoter underlined) and reverse primer 5'-AATAGAGCTTCCTTTAATTG.

Six mutations were introduced into the acceptor to identify regions preferred for
transfer. The primers used to introduce the mutations are listed in Table 3.

For the second set, the donor was derived from bases 2044-2463 of pNL4-3. The
primers used were 5'-

GATTTAGGTGACACTATAGTCAAAAGAAGGACACCAAATGAAAG (forward)

and 5'-CTTTATGTCCGCAGATTTCTATG (reverse). The acceptor was taken from
region 2014-2443 of the same plasmid and primers used to generate it were 5'

GATTTAGGTGACACTATAGAGGAAATAGGGCTGTTGGAAATG (forward)

and 5'-ATGAGTATCTGATCATACTG (reverse). The acceptor molecule contained
7 mutations incorporated using the primers listed in Table 3. The Multi Site Directed
Mutagenesis Kit from Stratagene was used to create all the above mentioned
mutations and the manufacturer's protocol was used to carry out mutagenesis and
transformations. XL10Gold cells were transformed as per the protocol. White
colonies were selected and their plasmids amplified and purified

Acceptor-Donor Set 1:

Primer	Primer sequence (5'-3')	Primer location on pNL4-3	Position of mutation on pNL4-3
1	taggaaccaaagaaagactg cta agtgttcaattgtggc	1938-1977	1958
2	tgcagggcccctaggaatagg g ctgttgaaatgtgg	2001-2039	2020
3	ccaaatgaaagattgtactgag t gacaggctaatttttaggg	2052-2094	2074
4	ccagggaaatcttcagag t agaccagagccaacagcccc	2122-2162	2142
5	gcttcaggtttggggaagagataacaactccctctcagaagc	2174-2215	2195
6	ccttagctccctcagatcattctttggcagcgaccctcg	2242-2283	2263

Acceptor-Donor Set 2:

Primer	Primer sequence (5'-3')	Primer location on pNL4-3	Position of mutation on pNL4-3
1	ccaaatgaaagattgtactgag t gacaggctaatttttaggg	2052-2094	2074
2	ccagggaaatcttcagag t agaccagagccaacagcccc	2122-2162	2142
3	gcttcaggtttggggaagagataacaactccctctcagaagc	2174-2215	2195
4	ccttagctccctcagatcattctttggcagcgaccctcg	2242-2283	2263
5	caattaaaggaagctctataagatacaggagcagatgatac	2304-2344	2323
6	gaattgccaggaagatggataccaaaatgataggg	2360-2396	2380

Table 3: **List of primers used to generate mutations in acceptor RNAs** of both Acceptor-Donor sets. The mutations introduced are shown in bold and their positions on pNL4-3 are indicated.

using a Qiagen mini-prep kit. The DNA was sequenced using T7 promoter. Once the sequence was verified, donor and acceptor substrates for both sets were amplified by PCR reactions that included 25 cycles of 94°C for 1 m, 50°C for 1 m, and 72°C for 1 m, followed by 72°C for 5 m. The products were electrophoresed on 6% native polyacrylamide gels prepared as described (110), located by UV-shadowing and excised then eluted in 10 mM Tris-HCl (pH=8.0), 1 mM EDTA (pH=8.0). Material was recovered by ethanol precipitation.

Transcription to generate RNA substrates: Approximately 1/3rd of the material purified after PCR was used in an in vitro transcription reaction using SP6 RNA polymerase from Roche Diagnostics according to the manufacturer's protocol. The transcription reactions were treated with 0.4 units/ μ l of DNase I/RNase free enzyme for 20 m at 37°C to digest away the template DNA. The RNA was then purified using the Qiagen RNeasy Mini Kit. This material was quantified using UV spectroscopy.

5' end-labeling of primers: Fifty pmoles of each donor reverse primer was 5' end labeled with γ -³²P ATP using T4 polynucleotide kinase following the manufacturer's protocol. Unincorporated ATP was removed using G-25 sephadex spin columns. Labeled primers were stored at -20°C.

Hybridization of nucleic acids: Primers (5' end labeled with ³²P) and corresponding RNA donor templates were hybridized in 80 mM KCl, 1 mM DTT, and 50 mM Tris-HCl (pH=8.0) and 0.1 mM EDTA (pH=8.0) at a ratio of 2:1 (primer:template) in a

volume of 20 μ ls by heating to 80°C for 3 min then slow cooling at a rate of 2° per min to 30°C. Hybridized material was used immediately in assays.

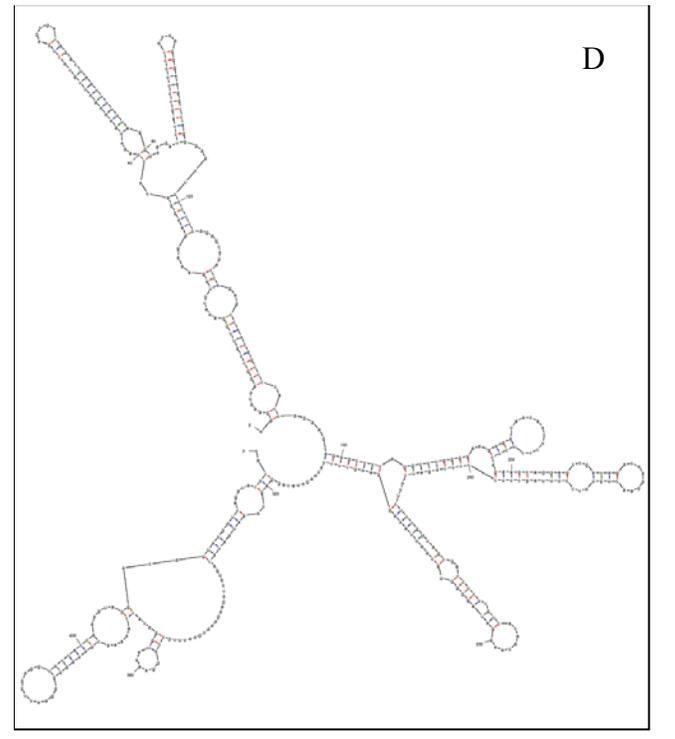
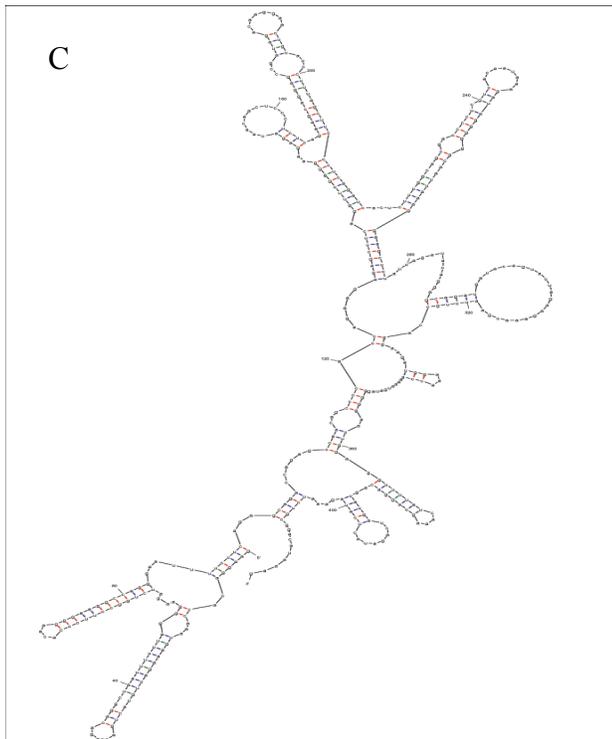
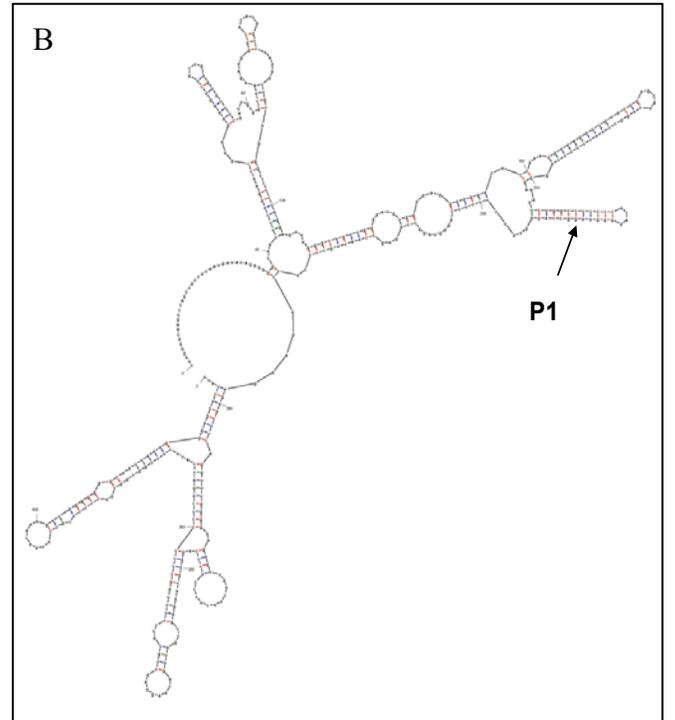
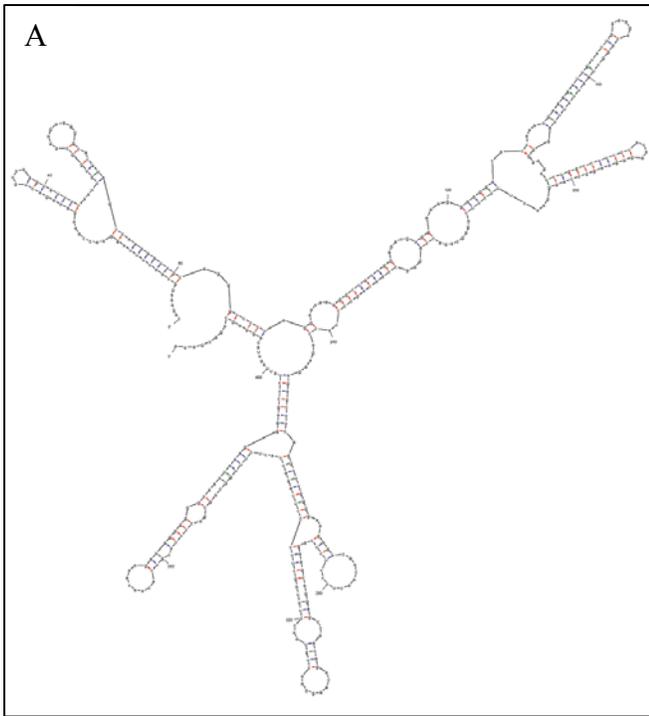
Time course reactions for strand transfers: Donor RNA and primer DNA hybrids (2 nM final concentration of RNA) were preincubated for 3 m at 37°C in presence of 2 nM acceptor RNA, NC (0, 1, 4 μ M) and MgCl₂ (1,6 mM) in a total volume of 42 μ ls. The reactions were initiated with 8 μ ls of RT at a concentration of 2.5 units/ μ l. The reaction was carried out in buffer containing the following components (final concentrations indicated): 50 mM Tris-HCl (pH=8.0), 0.1 mM EDTA (pH=8.0), 100 μ M dNTPs, 5 mM AMP (pH=7.0), 25 μ M ZnCl₂, 0.4 units/ μ L RNase inhibitor. The time points used for the reaction were 1, 2, 4, 8, 16, 32 and 64 m. At each time point, a 6 μ l aliquot was taken from the master reaction and stopped with a 4 μ l solution containing 25 mM EDTA (pH=8.0) and 26 ng DNase-free/RNase enzyme and allowed to digest at 37° C. Following this, proteinase K digestion was carried out with 2 μ l of the enzyme at a concentration of 2 mg/ml (enzyme buffer: 1.25% SDS, 15 mM EDTA (pH=8.0) and 10 mM Tris-HCl (pH=8.0)). The reaction was carried out at 55°C for 60 m. Lastly, 12 μ l of sample buffer (2X: 90% formamide, 10 mM EDTA, 0.25% each bromophenol blue and xylene cyanol) was added and the reactions were run on a 6% denaturing polyacrylamide gels (7M urea, 19:1 acrylamide:bisacrylamide) as described (110). Products were detected and quantification was performed using a BioRad model FX phosphoimager.

Amplification and sequencing of DNA transfer products: For each of the six conditions used (0 μM NC/1 mM Mg^{2+} , 0 μM NC/6 mM Mg^{2+} , 1 μM NC/1 mM Mg^{2+} , 1 μM NC/6 mM Mg^{2+} , 4 μM NC/1 mM Mg^{2+} , 4 μM NC/6 mM Mg^{2+}), a 32 m strand transfer reaction was carried out. The procedure was the same as mentioned above except that the reaction volume was 50 μl . The reactions were processed and then electrophoresed on a 6% denaturing polyacrylamide gel. Strand transfer products were detected by exposing the gel to a phosphoimager screen and were then excised, and eluted overnight in TE buffer (1 mM EDTA (pH=8.0) and 10 mM Tris-HCl (pH=8.0)). The eluate was filtered and precipitated in ethanol and then resuspended in 30 μl of RNase free deionized water. This was then used to amplify strand transfer products using PCR (PCR conditions as mentioned above) and the following primers: 5'-GATTTAGGTGACACTATAGCAAGTAACAAATCCAGCTAC and 5'-TATCATCTGCTCCTGTATCT. The products of PCR were cloned into the Strataclone vector and used to transform Strataclone SoloPack Competent cells using the manufacturer's protocol. White to pale blue colonies were picked for plasmid purification using the Qiagen miniprep kit. The plasmids were sequenced using T3 promoter.

3.3 Results

3.3.1 Generation of RNA Acceptors and Donors

The two sets of RNA acceptor and donor sequences used were made using primers listed in the previous section. The predicted structures of both sets of sequences were determined using mFold (132). Identical salt and temperature conditions were used to predict each structure. Mutations made on the acceptors were chosen in such a way that their structures would remain similar to their corresponding donor sequences (Fig. 15A-D). For Acceptor-Donor set 1 only small differences in the predicted lowest energy structures was observed. In contrast, set 2 showed more dissimilarity although all strong stem-loop structures were essentially conserved between the donor and acceptor in the region of homology. Each acceptor donor pair was chosen such that the product of strand transfer would be 30 bases longer than a donor directed DNA product due to an additional 30 base extension on the acceptor. In addition, a 5 nucleotide stretch was added to the 5' end of each donor RNA that was not homologous to the acceptor template. This was done to avoid transfer of a fully donor directed product. Therefore, this system tests for internal strand transfer (see Fig. 15E). These changes and the additional mutations in the acceptor template are necessary for detection of strand transfer products in the system. However, they can contribute to small structural differences between the donor and acceptor.



E.

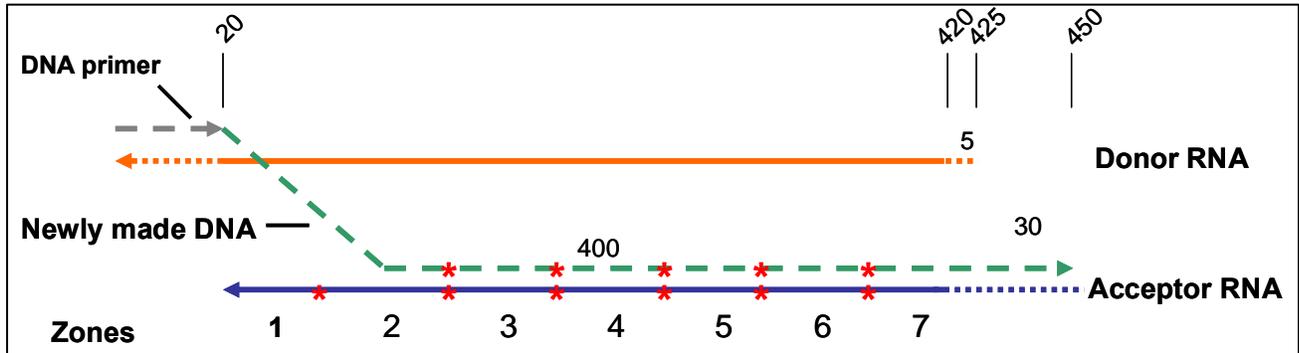
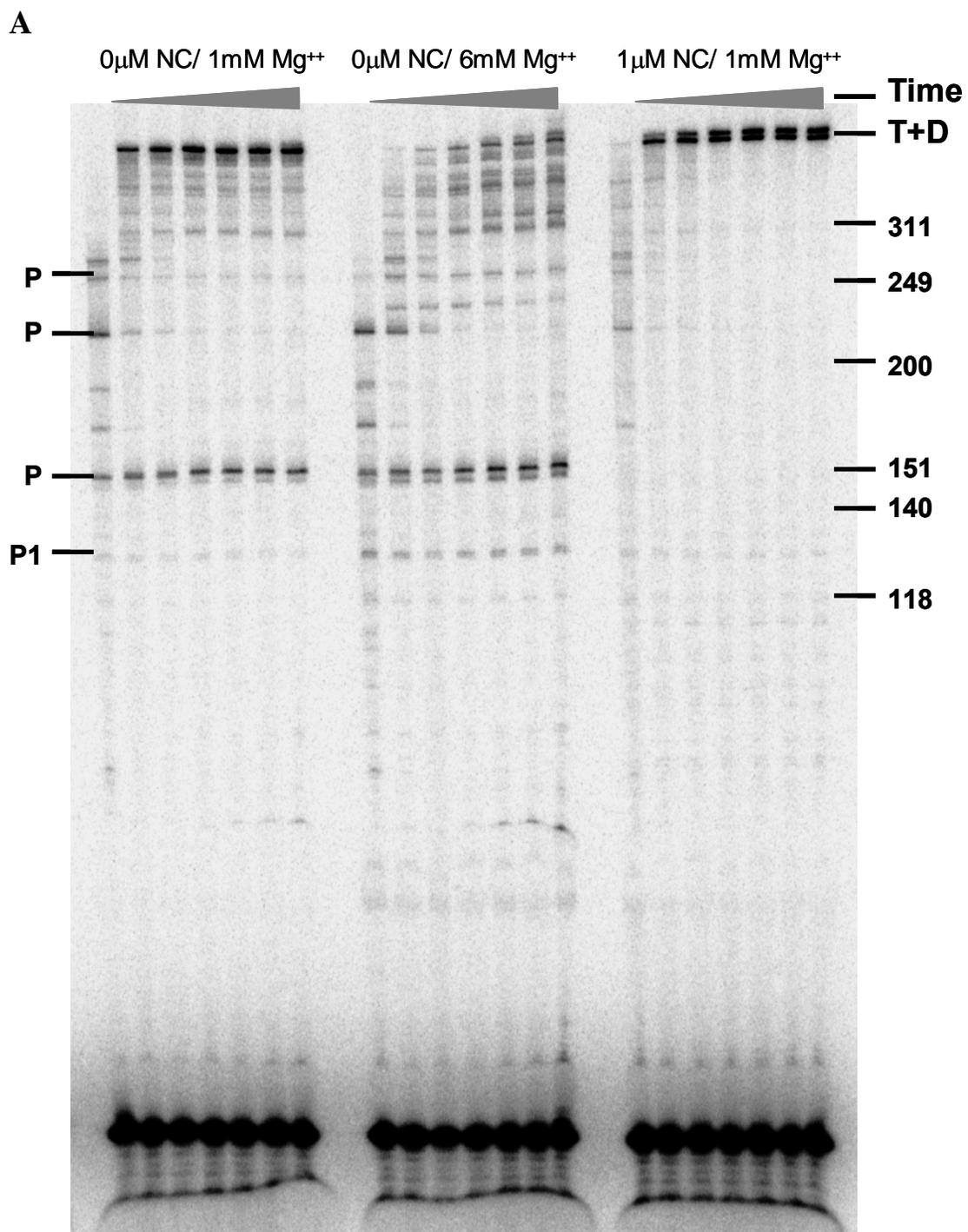


Figure 15: A-D- **Structures of Donor and Acceptor RNAs** of Set 1 (A and B respectively) and Set 2 (C and D respectively). Predictions were made using *mfold* software at 37 °C and 1 M salt (132)

E- Schematic representation of the Acceptor-Donor system used to study recombination in vitro. Solid lines represent the 400 nt region of homology on the RNA templates, while dotted lines show the excess regions used to differentiate between transfer and donor directed products (see subsection 3.3.1). Dashed lines indicate DNA primer and the product made from it. Mutations are indicated with a red asterisk. The zones created by incorporating mutations are numbered in the direction of DNA synthesis. Crossing over in any zone would cause the nascent DNA to exhibit all downstream mutations.

3.3.2 Effect of Mg^{2+} and NC on RT synthesis

To study the effect of NC and Mg^{2+} on strand transfer on Set 1, strand transfer time course experiments were carried out (as described in Methods) using 0, 1 and 4 μM NC. For each concentration of NC, two different amounts of Mg^{2+} were used (1 and 6 mM). Aliquots were removed from the strand transfer reaction between 1 to 64 minutes and analyzed by gel electrophoresis. The pattern seen for donor/acceptor set I is shown in Figs 16A and 16B. In the absence of NC, pausing by RT was pronounced (some prominent pause site (P1, P2 etc.) are labeled on the gels in Figs 16A and B). Also, a marked pause site was seen at ~ 120 bases (P1). This region corresponds to a stem loop structure in the donor RNA (Fig. 15A-D). Increasing the $[Mg^{2+}]$ from 1 to 6 mM significantly increased pausing and decreased the level of fully extended products consistent with Mg^{2+} stabilizing secondary structures in the RNA template. When 1 μM NC was included in the reactions the amount of pausing significantly decreased and the level of fully extended products increased. With 1 mM Mg^{2+} very little RT pausing was observed. Pausing was considerably more prominent when 6 mM Mg^{2+} was used. Still more pausing was seen under this condition in the absence of NC. When 4 μM NC was used a clear inhibition in the level of extended primer was noted, especially with 1 mM Mg^{2+} . Inhibition of primer extension at high NC concentrations has been observed previously in our lab ((8), see Discussion). Pausing at 1 mM Mg^{2+} was comparable to the 1 mM Mg^{2+} /1 μM NC experiment (Fig. 16). However, in contrast to the 1 μM NC results, increasing $[Mg^{2+}]$ to 6 mM did not significantly change the level of pausing. Overall the results show that Mg^{2+} increasing pausing



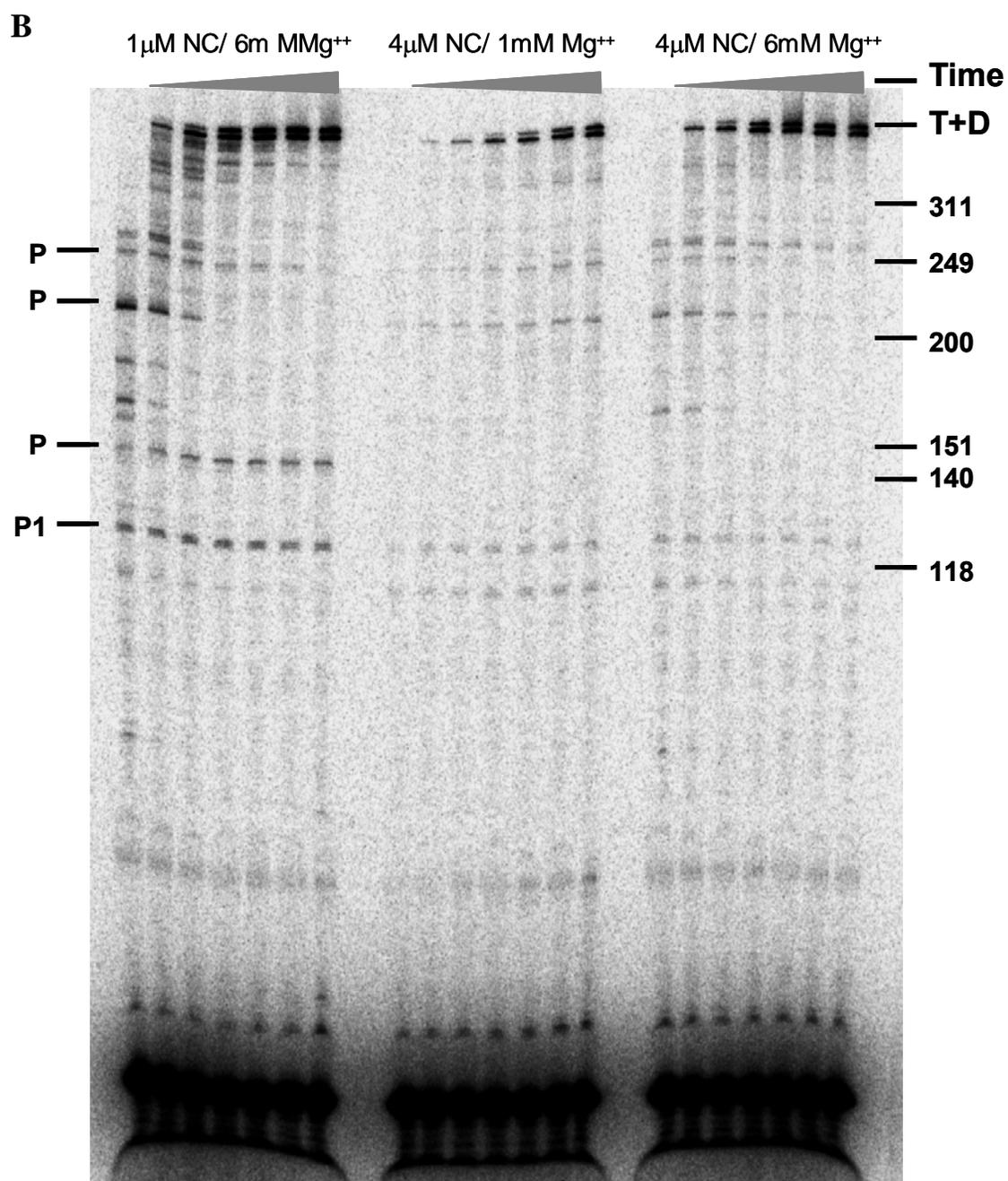


Figure 16: A and B- **Extension by RT in Acceptor-Donor Set 1**. Extension of the primer under different [NC] and [Mg²⁺] is shown as indicated. The experiment was carried out for 1-64 min. the major pause sites are indicated (P1 and P2). The total extended product (T+D) includes both transfer (T) and donor directed (D) full length DNAs.

while NC decreases pausing as expected. Of the two, the NC effect is greater resulting in strong decrease in pausing even with high Mg^{2+} .

Previous results have shown that internal strand transfer is highly stimulated by NC protein and pause sites can serve as focal points for transfer events. Therefore the finding that pausing patterns are altered by different Mg^{2+} and NC concentrations suggested that transfer may also be affected. In order to clearly separate and quantify 425 nt donor directed and 450 nt strand transfer products reactions were carried out for 32 min and run for an extended period of time on denaturing gels (see Fig. 17). Products were measured using a phosphoimager. Strand transfer levels are expressed as % transfer efficiency which is the amount of transfer product (T) divided the total of transfer plus full length donor directed (D) products and multiplied by 100 $((T/(D + T)) \times 100)$. These values are listed in Table 4 for the various conditions. For experiments with 1 mM Mg^{2+} , transfer efficiency increased ~3-fold with 1 μ M vs. no NC then some decrease was observed with 4 μ M compared to 1 μ M. With 6 mM NC transfer approximately doubled from 0 to 1 μ M NC then stayed constant with 4 μ M NC. For each NC condition an increase in transfer efficiency was observed when 6 vs. 1 mM Mg^{2+} was used. This was especially notable for the no NC and 4 μ M NC conditions. Overall NC enhanced strand transfer as expected. Transfer was also enhanced with Mg^{2+} . The latter effect is consistent with more pausing occurring in the presence of higher Mg^{2+} .

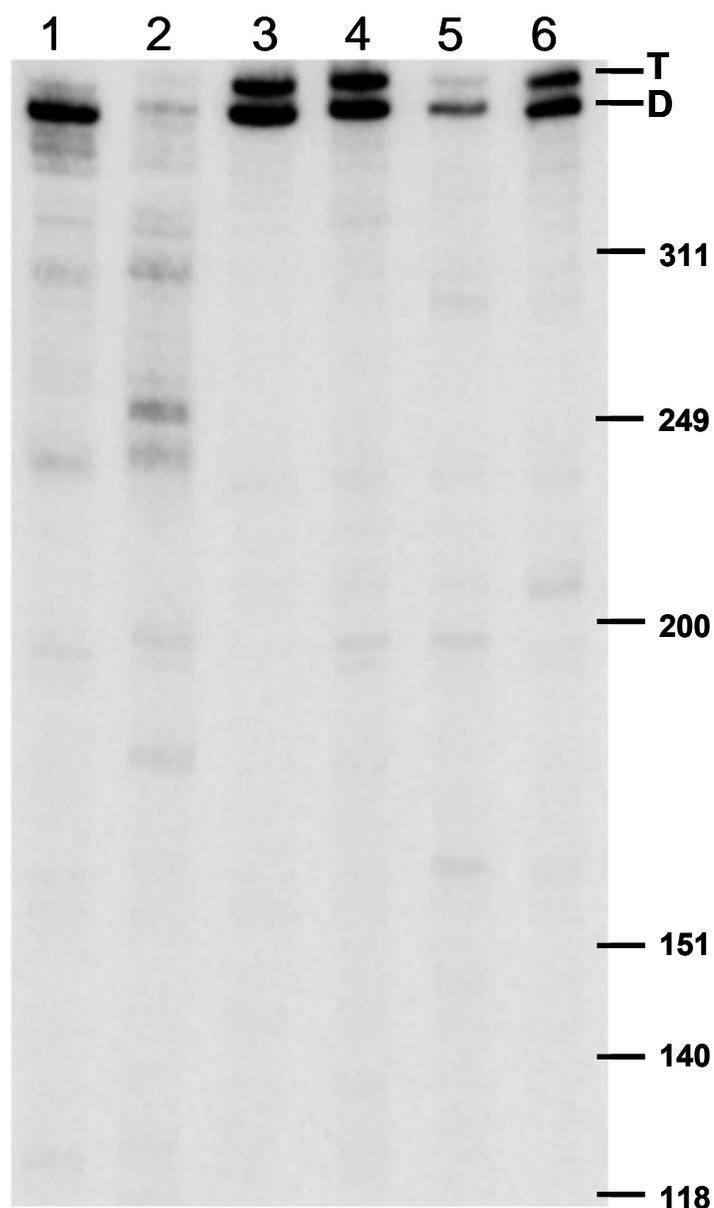


Figure 17: **An experiment to calculate strand transfer efficiency Acceptor-Donor Set 1.** The protocol used was the same as that for the assay shown in Fig. 16. The autoradiogram here shows reactions under different conditions allowed to extend for 32 min. The transfer (T) and donor directed (D) products were separated and quantified using phosphor imager analysis. Reactions in lanes 1-6 were carried out at 0 μM NC/1 mM Mg^{2+} , 0 μM NC/6 mM Mg^{2+} , 1 μM NC/1 mM Mg^{2+} , 1 μM NC/6 mM Mg^{2+} , 4 μM NC/1 mM Mg^{2+} , and 4 μM NC/6 mM Mg^{2+} respectively. The numbers obtained are listed in Table 4

Reaction condition	Efficiency of strand transfer
0 μM NC/1 mM Mg^{2+}	12.0 \pm 6.0
0 μM NC/6 mM Mg^{2+}	27.0 \pm 14.0
1 μM NC/1 mM Mg^{2+}	41.0 \pm 11.0
1 μM NC/6 mM Mg^{2+}	50.0 \pm 6.0
4 μM NC/1 mM Mg^{2+}	27.0 \pm 10.0
4 μM NC/6 mM Mg^{2+}	48.0 \pm 8.0

Table 4: **Strand transfer efficiency for Acceptor-Donor Set 1 under different reaction conditions.** Transfer (T) and donor directed full length (D) products from experiments such as the one shown in Fig. 17 were quantified and the efficiency of transfer was calculated as - Efficiency = $[T/(T+D)]*100$. The experiments were done in triplicate.

3.3.3 Frequency of hot-spots in Acceptor-Donor Set I in vitro

To assess the frequency of strand transfer at different positions of the 400 base homology region of Set I, an acceptor with six mutations was used (see Fig. 15E). The mutations were approximately 60 bases apart, thus marking 7 zones in which the rate of recombination was studied. The mutations were created using primers listed in Table 3. Strand transfer reactions were carried out for 32 min. at each of the given NC and Mg^{2+} condition and the resultant strand transfer products were isolated from gels then amplified and cloned into a Stratagene vector and were then sequenced after plasmid purification (see Methods). At least 30 plasmids of each reaction condition were sequenced to see where transfers occurred. The position of crossover from the donor to the acceptor was scored based on where the sequenced DNA product had acquired one of the six mutations on the acceptor. It was assumed that the crossover had occurred between this mutation and the previous mutation nearer the 3' end of the acceptor.

Calculations: the frequency for recombination in each region was calculated using the equation below. Earlier studies in the field have used the percentage of total transfers occurring in each region to estimate efficiency in a given region. A different approach is used here because of the comparatively high efficiency of transfer overall. The reasoning is as follows: For any given region, all extended DNAs that transferred in prior regions are no longer eligible for transfer in that region. This implies that at later regions, fewer DNAs are available for transfer. Therefore estimating transfer in each region as a percentage of the total number of transfer events that were sequenced,

would not represent the actual probability of transfer in that zone. Hence we used an equation where the efficiency of transfer was calculated based on what fraction of sequences that could possibly transfer in a particular region actually did transfer in that region.

The analysis assumes that each DNA extended to the end of the donor (D^*) or end of the acceptor (T^*) represents a potential recombination event. For calculations purposes, the total number of sequenced transfer products under each condition was set equal to T^* . The ratio of transfer (T) to transfer (T) + donor directed (D) products for a given condition (see Table 4) was used to calculate total potential recombination events T^*+D^* ($T^*+D^*= T^*/[T/(T+D)]$).

Once the total potential recombination events were determined (T^*+D^*), the number of transfer events in each region (E_n) was divided by the potential recombination events for that region $(T^*+D^*)_n$ to calculate what fraction of potential recombination products actually transferred in that zone (F_n). The potential recombination events in each region was computed by subtracting recombination events occurring in all zones prior to the one in question from the total potential recombination events (T^*+D^*). Fractions calculated for each zone were then divided by the sum for all 7 zone and multiplied by 100 to give the percentage of transfer in each zone ($[F_n/(F_1+F_2+\dots+F_7)] \times 100$). In comparison to simply dividing the number of recombination events in a given zone by the total number of transfer events sequenced ($E_n/T^* \times 100$), the above interpretation result in a small skewing of the data toward latter regions.

$$\text{Frequency of transfer in a given zone} = \frac{F_n}{F_1+F_2+\dots+F_7} \times 100$$

$$\text{Where, } F_n = \frac{E_n}{(T^*+D^*) - \text{Transfer events in all previous zones}} \times 100$$

$$\text{Where, } E_n = \text{Transfer events in a particular zone}$$

$$\text{And, } (T^*+D^*) = T^*/[T/(T+D)]$$

Where, T*= Total potential transfer products
D*=Total potential donor directed products
T = Transfer products (quantified from assay)
D = Donor directed products (quantified from assay)

The results obtained for each reaction are shown in Fig. 18.

In the absence of NC, almost all of the strand transfer events occurred towards the end of the donor template indicating reduced transfer events in the preceding regions. This is consistent with the fact that in the absence of NC, a greater degree of homology is usually required for transfers to occur. At 0 mM NC, increasing the amount of Mg²⁺ in the reaction did not significantly change the efficiency of transfers in each zone. As the amount of NC in the reaction was increased, a shift was observed in the frequency of transfer toward the middle and beginning of the template, closer to where RT synthesis initiated. At 1 μM NC transfer shifted towards the central zones while regions 5 and 6 showed the greatest transfer for 1 and 6 mM Mg²⁺

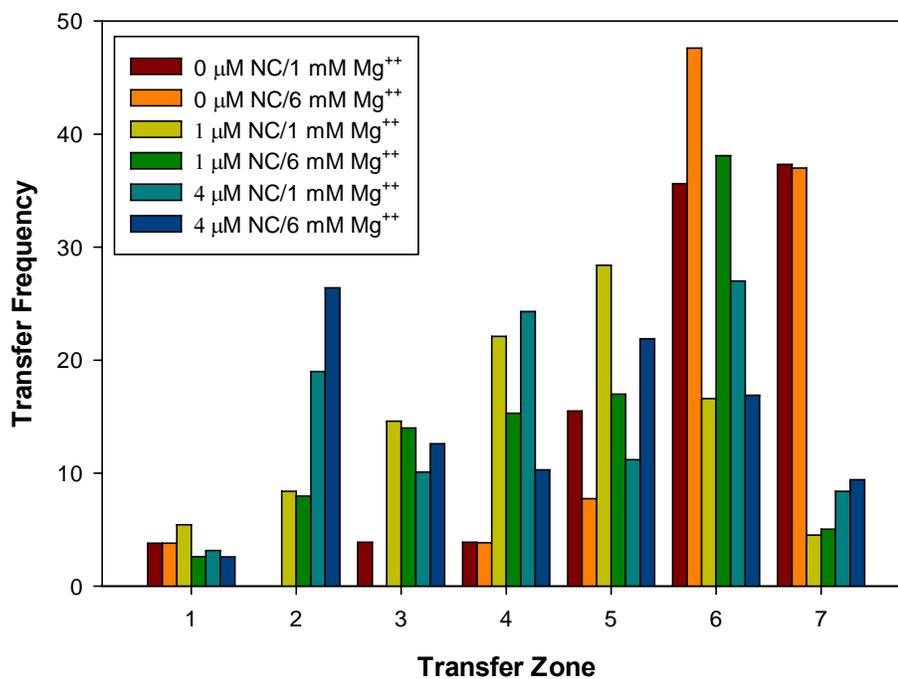


Figure 18: **Frequency of transfer in each zone of Acceptor-Donor Set 1 in vitro.** Transfer zones were marked as shown in Fig. 15E. A 32 min reaction was carried out for each reaction condition and run on a denaturing gel. Transfer products were excised and cloned and sequenced. The location of transfer was determined based on the presence of acceptor mutations. Transfer frequencies for each condition were based on the calculations described above.

respectively. Under this condition, transfer efficiency for 1 or 6 mM Mg^{2+} essentially mimicked each other in all zones except 5 and 6. This shift was more prominent at 4 μ M NC where strand transfer was essentially distributed along the entire template except for regions 1 and 7 where transfer was low. The low level of transfer under all conditions in zone 1 is predictable since the region of complementarity between the acceptor and nascent DNA in this zone is small only reaching 61 maximal bases at the end of the zone. The distributing of transfers was also similar at 1 and 6 mM Mg^{2+} and differed most significantly in regions 4 and 5. The increase in transfer at earlier zones on the template is consistent with transfer being able to occur with less homology in the presence of higher amounts of NC (36). These results were then compared to strand transfer patterns obtained in cell culture.

3.3.4 Frequency of hot-spots in Acceptor-Donor Set I in cell culture

This work will be carried out by Dr. Negroni's lab using a method that has already appeared in literature (55, 56). Briefly, helper cells are used to package two different viral constructs and produce viral progeny (both homo- and heterozygous) capable of infecting MT-4 cells and producing viral DNA, but not new infectious viruses. Low molecular weight DNA is then purified from the cells post-infection, amplified and processed as outlined in Fig. 14. Recombination events can be estimated because some of the viruses made by the transfected helper cells will carry genomic RNA from both constructs (heterozygous). The two constructs are characterized by the presence of a functional *lac* gene (*lac*⁺) on one and an *E. coli* *malT* gene on the other (*lac*⁻). The *lac*⁻ vector also has an additional BamHI restriction

site. Both have deletions in the U3 region and a deletion of the “*flap*” region to limit nuclear import. The constructs share a region of homology adjacent to the *lac* or *malT* genes. In our experiments, one construct will have the wild type *gag-pol* 400 nt homologous region described above for the in vitro work while the other will have the version with the 6 mutations. Crossovers from the *lac*⁻ to the *lac*⁺ genomes that occur in the homologous region will pick up the functional *lac* gene (assuming they don’t cross back to *lac*⁻, an event which can be estimated from the results). A strategy using PCR and selective restriction digests will be employed to isolate the *lac/malT*-homology region sequence that contains the BamHI site. This DNA will be cloned into a plasmid vector that can be used for blue/white screening in a suitable bacterial host. BamHI selection will limit the analysis to the parental *lac*⁻ or recombinant *lac*⁺ products producing white and blue colonies, respectively. The frequency of recombination will be estimated using various controls and the proportion of blue to total colonies, and sequencing of plasmids prepared from blue colonies will be used to map crossover sites. Several labs (Pathak and Hu (69), Dougherty (131), Telesnitsky (96), and Levy (89) labs for example) have similar systems that differ either by the various marker proteins used or whether the genomes for the new infectious viruses originate from virus-derived inserted proviruses or plasmid RNA, as is the case in Dr. Negroni’s system.

The results obtained from such an analysis of Acceptor-Donor Set I is shown in Fig. 19. Of the 25 sequences analyzed, the majority of crossovers took place in the first zone (44%). The rest of the zones showed small amounts of transfer (ranging from 8-16%) except for zone 4 which did not have any transfer events. As in the in

vitro experiments, transfer was spread out over the entire 400 nt region except for zones 1 and 4. However, the cell culture results did not match the recombination pattern of any one condition used in vitro. Zone 4 which was earlier found to contain a strong hotspot (36) (in vitro) in the context of a smaller region of homology, did not show any strand transfer events. This could be due to the small number of sequences analyzed in cell culture. Only 14 total recombination events were detected for zones 2-7 combined. This number is clearly too low to attempt to make a correlation between the in vitro and cellular results (see Discussion). Zone 1 which was a poor hotspot in vitro for the reasons noted above, was a major recombination break point in culture. While this could have resulted from a strong predicted secondary structure in this zone (Fig. 15 A-D), it is also possible that the region of homology preceding this zone in the cell culture constructs may have facilitated transfer. Zone 1 is flanked by a modified U3 region in the culture experiments which essentially results in several hundred homologous bases between the donor and acceptor constructs that precede zone 1. This high degree of homology may have driven recombination in this zone. But it is also possible that the zone contains a natural hotspot that simply cannot be detected in vitro because of low homology in this zone. To test this, we designed Acceptor and Donor Set 2 where zone 1 of Set I was now in the middle of the region of homology.

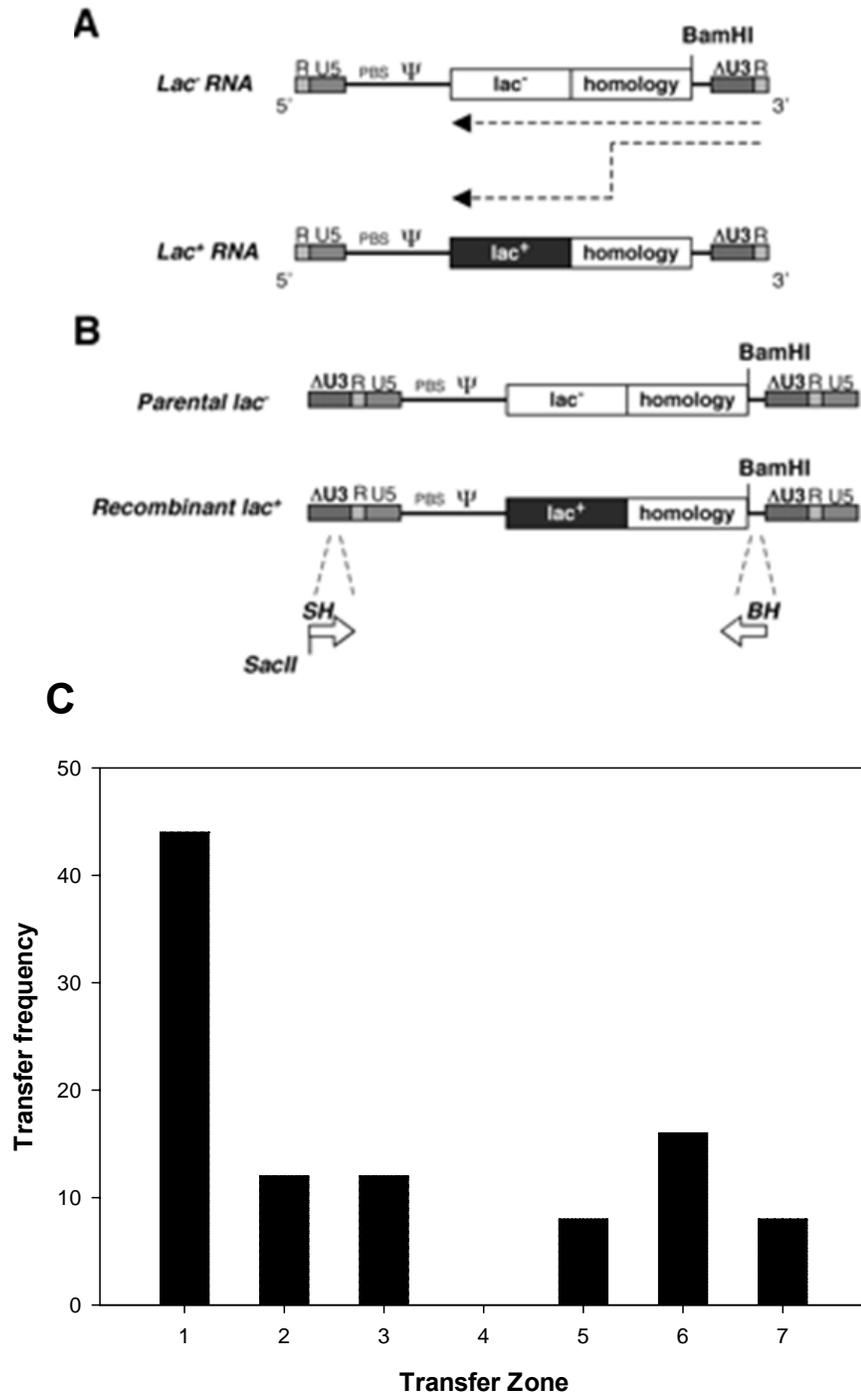


Figure 19: A and B- Cell culture system used by Dr. Negroni to study recombination in a single replication cycle. (See text for details).

C- Transfer frequencies obtained from cell culture assays. A total of 25 samples were analyzed.

3.3.5 In vitro results for frequency of hot-spots in Acceptor-Donor Set II:

Results obtained from sequencing recombinants from the second set of Acceptor and Donor are shown in Fig. 20. Like Set I, these results show that a build-up of homology is required for transfers to occur in the absence of NC. The addition of NC to the reaction facilitates transfer at earlier locations of the region of homology. At the highest concentration of NC used (4 μM), transfers significantly shifted to the initial zones. Mg^{2+} , however, did not have a consistent effect on where transfers occurred. In a time course reaction using the same acceptor-donor pair, NC was seen to increase transfers and reduce pausing, while Mg^{2+} was seen to increase both pausing and total strand transfer (data not shown). This is consistent with results obtained from a time course assay with Acceptor-Donor pair of Set 1 (Fig. 16). This suggests that the major contributor to the location of strand transfer is NC concentration and that while Mg^{2+} may affect pausing, its effect on recombination break points may not be as profound as NC's.

Interestingly, zone 3 of this set (which corresponds to the major hotspot found in cell culture with Set I), was not a major recombination hotspot. This suggests that the concentration of transfers seen in this region in cell culture may have been due to the homology between the constructs in the prior region. Our collaborator is now analyzing Set II sequences for recombination hotspots.

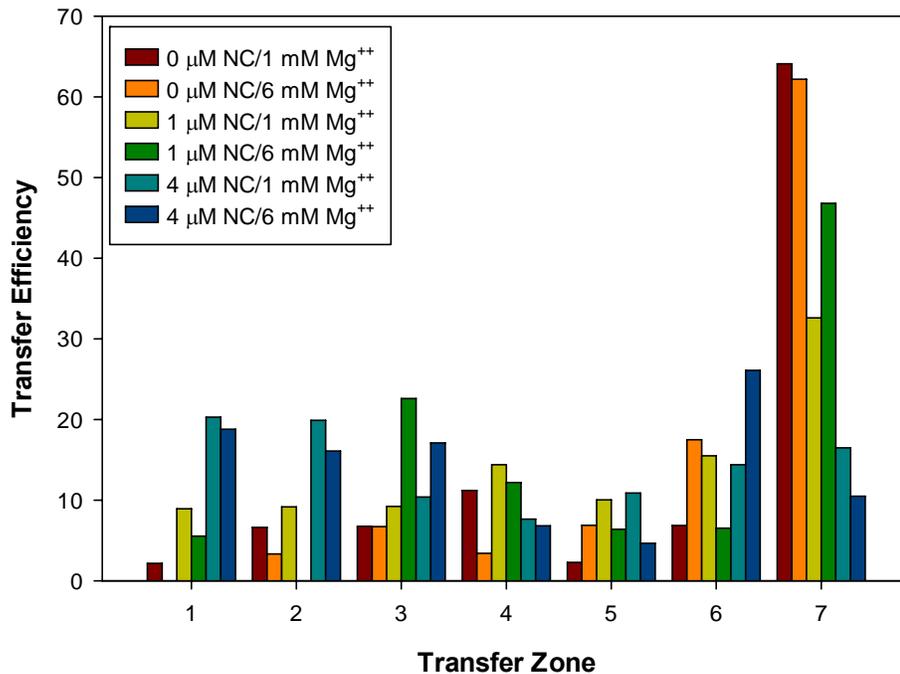


Figure 20: **Frequency of transfer in each zone of Acceptor-Donor Set 2 in vitro.** Transfer zones were marked as shown in Fig. 15E. A 32 min reaction was carried out for each reaction condition and run on a denaturing gel. Transfer products were excised and cloned and sequenced. The location of transfer was determined based on the presence of acceptor mutations. Transfer frequencies for each condition were based on the calculations described earlier.

3.4 Discussion

In this report I look at the factors affecting recombination in vitro and if and under what conditions these results can be correlated to what happens in the cell. Results of the experiments performed here suggest that NC and Mg^{2+} concentrations in reconstituted reactions affect the location of strand transfer and the extent of pausing during extension. NC has been studied extensively with regard to strand transfer in vitro. The first obligatory strand jump occurs only in the presence of NC in the Gag precursor (31). Internal strand transfers are also facilitated by NC. NC has been shown to increase the RNase H activity of RT as well as remove RNA cleavage products generated during DNA synthesis and promote annealing (88). Together, this could result in an increase in degradation of the donor strand and subsequent association to the acceptor resulting in strand transfer. NC also melts secondary structures (16) thus decreasing pausing during extension. The results obtained here show that NC does have a strong effect on the profile of strand transfer and the extent to which strand transfer occurs. In the presence of NC, recombination occurred towards the beginning of the region of homology suggesting that a build up of homology wasn't as necessary in the presence of NC. As expected, NC also reduced pausing over time possibly by melting secondary structures thereby allowing RT to traverse the template with less resistance. NC caused transfer events to occur at the beginning of the region of homology in a dose dependent manner. The shift toward earlier regions was evident between 1 and 0 μ M NC then again between 4 and 1 μ M NC. It would be interesting to test some NC concentrations between 1 and 4 μ M and

perhaps even higher concentrations to determine if the shift effect is saturated at 4 μM .

Magnesium ion concentration can affect the structure of nucleic acids and the binding and enzymatic activities of proteins during strand transfer (see chapter introduction). The results we obtained showed that while recombination was affected by the concentration of Mg^{2+} , the effect wasn't as pronounced as that of NC. However, Mg^{2+} did alter the extent of pausing during extension. An increase in the concentration of Mg^{2+} in the reaction resulted in an increase in pausing and the efficiency of strand transfer products (Fig. 16 A and B). This is in keeping with the fact that Mg^{2+} increases the formation of secondary structures, thus increasing pausing. It was interesting that the effect of Mg^{2+} was essentially masked at the highest NC concentration where no significant increase in pausing was observed with higher Mg^{2+} . At the lower NC concentration the effect of Mg^{2+} was only partially masked. A clear increase in pausing occurred with 6 vs. 1 mM Mg^{2+} however, pausing was still considerably less than in the absence of NC. In the virion the concentration of NC is predicted to be ~ 10 mM, an amount not possible to analyze in vitro due to aggregation. According to our results, under these conditions Mg^{2+} is unlikely to play a significant role in pausing and would have minimal effect on RT's traversing the template. It is important to point out however, that although HIV replication presumably begins in a core-like structure, how the structure changes during replication is unclear. It is possible that NC could be released as the core dissipates. Therefore it is unclear what the concentration of NC is throughout the entire replication process.

Interesting, no apparent decrease in RT activity was observed with 1 vs. 6 mM Mg^{2+} under any condition. In fact synthesis was actually more efficient with lower Mg^{2+} as greater proportion of extended primers reached the end of the templates. This was unexpected since RT shows highest polymerization and RNase H activity rates between 4-8 mM with only about $\frac{1}{2}$ maximal activity at 1 mM Mg^{2+} (28, 117). This could possibly be due to the reduction in secondary structures at lower concentrations of Mg^{2+} which allows RT to read the template with lesser resistance though RT activity at this Mg^{2+} concentration is lower.

With regard to the frequency of transfer in the zones marked by mutations on the acceptor, Mg^{2+} did not display as marked an effect on where transfer occurs as did NC. For each different concentration of NC, the effect of increasing Mg^{2+} concentrations did not consistently increase or decrease the extent of transfers in a particular zone. This suggests that though Mg^{2+} can increase the efficiency of strand transfer at a given NC concentration, this effect is not dependent on any single location on the template.

Results presented here show that NC and Mg^{2+} both affect recombination albeit to different extents. Overall, NC was found to have a greater effect on the extent and location of strand transfer than Mg^{2+} . Increasing the concentration of the cation in the reaction caused an increase in pausing during extension over the template RNA. This would suggest that while Mg^{2+} in cells can affect the polymerization rate of RT and the structure of the RNA, NC has a more significant effect on the position and level of recombination.

The analysis of the region of homology in cell culture gave results that did not correspond to those obtained in vitro. This could be due to a number of reasons. For example, only 25 sequences were analyzed in the cell culture studies. Of these 11 transfers occurred in the first zone (which may be due to the region of homology ahead of the zone of interest, see Results). Hence, only 14 transfer events were used to analyze the remaining 6 zones. Analyzing a larger number of sequences in cells would give a more statistically significant estimate of transfer points. Also, the inherent differences in systems may affect the results obtained from the region of interest. For example, in in vitro assays, the only homology seen is within the region of interest, however, the cell culture constructs require homologous regions flanking the region tested. This could affect the outcome such that a moderate recombination hotspot could now become more prone to transfers due to the build up of homology in the extraneous regions.

Also of interest is the fact that transfer zone 4 in Acceptor Donor Set I contains a previously defined hotspot (36). This region showed moderate transfer in the experiments in vitro, while no transfers occurred in this region in cell culture studies. This could be because the zones analyzed here are ~ 60 bases long and it is possible that all the crossovers observed in one region could arise from the same nucleotide location, while the large number of transfers seen in another zone might be dispersed over the 60 nt region. This could be addressed by looking at smaller zones for transfers, with mutations closer together as was done in the previous study identifying the hotspot (36). However, this has to be done carefully, since introducing

mutations too close to each other would lead to a decrease in homology between the donor and the acceptor in the region of interest which may influence recombination.

Chapter 4: General Discussion

This aim of this dissertation was to analyze factors of both viral and host origin influencing HIV-1 replication *in vitro*. Chapter 2 of this dissertation looks at the possible role of NC in priming with HIV RT. NC is an extensively studied HIV protein and is important for almost all the steps involved in reverse transcription (for a review see (88)). Hence, NC's contribution was a major focus of our study of primer extension by RT. As has been described in Chapter 2, it has been shown for the first time that NC affects priming of PPT and non-PPT primers differentially, suggesting a possible role during plus strand priming in the viral core (73). This *in vitro* effect of NC has been validated in a recent, unpublished work of another group that uses a different approach (Dr. J. Levin, personal communication). Results obtained here show that the helix destabilization activity of NC and the competition between RT and NC to bind in an extension favorable mode to primers are responsible for NC's effect on plus strand priming. What we observe is that RNA primers can be generated and used by RT during reverse transcription and in the presence of NC, non-PPT primer extension is strongly inhibited (Figs. 13 and 14). According to the model we propose here, NC does this in two ways. Firstly, for short non-PPT RNA primers generated by RT, the helix destabilization activity of NC causes the dissociation of the primer template hybrid thus inhibiting extension by RT (Fig. 9). Secondly, for longer hybrids, where destabilization is less feasible, there is a competition between RT and NC to bind to the 3' end of the primer RNA (Fig. 10). This is because the preferred binding orientation of RT on an RNA primer is with its

polymerase domain to the 5' end (40), thus making binding in an extension mode relatively weak. This coupled with NC's ability to bind nucleic acids makes extension all the more difficult. Both of the above effects are not seen with the PPT because: a) the PPT is relatively resistant to internal degradation by the RNase H activity of RT, hence it cannot be easily dissociated from the template and b) unlike all other RNA primers, the PPT orients RT in such a way that its polymerase domain is directed to the 3' end of PPT (40).

The extent to which each of these effects influences the final outcome is not easy to determine due to the close association of both these properties of NC. While the use of NC mutants with normal binding activity but lowered unwinding activity is possible, mutants with the opposite properties have not yet been clearly established. It is difficult to imagine a mutant that binds poorly and yet causes helix destabilization at wild type levels. In the experiments described here, primer extension in the absence of RNase H activity of RT (and the consequent generation of short hybrids) was used to overcome this limitation. Another possible method to directly assess the effect of binding of both NC and RT during extension would be to use trap assays where the amount of synthesis during a single round of RT extension can be studied in the presence of wild type and mutant NCs. This would help us assess how well RT associates with a primer (PPT and non-PPT) in an extension favorable mode in the presence and absence of NC. Though this was attempted in this report, the low affinity of RT for all RNAs including the PPT made K_d measurements difficult. Increasing the substrate in the reaction and studying binding and extension at very short time points using a rapid quench system could perhaps yield more data in this

regard. The orientation of RT on RNA and DNA in different primer-template combinations have already been shown biochemically (40). A recent study using single molecule fluorescence resonance energy transfer assays has also demonstrated the dynamics of RT binding to primers (4). The use of more sensitive methods like this to study RT binding affinity and dynamics in presence of NC (wt and mutants), could further elucidate the relation between RT binding, orientation of RT binding and interference by NC on extension of PPT and non-PPT RNA primers.

One the most interesting and significant assays in this study was the use of a long RNA template to study the effect of NC on the generation and extension of primers by RT (Fig. 12). This assay clearly showed that RT can create and use primers other than the PPT during extension over a long RNA template as would occur during minus and plus strand synthesis respectively (Fig. 13). In the system used here, a 430 nt stretch of RNA from the *gag-pol* region of the HIV genome was used to study priming and the same region with the PPT sequence inserted in the center was used as a control. The preference of RT for PPT both in the presence and absence of NC was marked. Also of significance was the observation that other RNA fragments were used for priming and though the presence of NC did cause a decrease in their extension, it did not fully eliminate their use. Some of these regions remained constant across different reaction conditions suggesting that they were intrinsically more favored for extension. It would be very interesting to test these non-PPT RNA primers for similarities in sequence and structure to the PPT. Studying the region that flanks the PPT in the virus could also give us some insight into the preferential use of PPT as a primer. An approach similar to the one described above but using the natural

location of PPT might show if the RNA in this region has more “primable” sequences and whether NC affects their extension as well. An RNA template derived from the same region but lacking the PPT sequence would be a good control in this case. As an extension of this idea, PPT containing regions from different retroviruses can be studied for priming by their RTs in presence and absence of their cognate NCs. This would be especially helpful in some cases like that of avian retroviruses that quite frequently use sequences other than their PPT for plus strand priming. A correlation between the specificity of primer use shown by each and the binding and inhibition patterns shown by their corresponding RTs and NCs would be valuable in further understanding the role of NC in priming.

As with any in vitro study, the significance of an observed effect cannot be easily translated into what actually occurs in the cell. Is non-PPT priming inhibition by NC required for replication of HIV in the host? This question cannot be answered without extensive study in both the test tube and cell culture. However, the fact that HIV has a strong preference for PPT as its plus strand primer and the observation that in the presence of a completely randomized PPT, it quickly reverts back to the original sequence suggests that this sequence may be evolutionarily beneficial to the virus (91). Indeed, it is plausible that RT and PPT have co-evolved for tighter binding as compared to other RNA primers (41). This may be important for generation of the correct ends for integration (as the PPT/U3 junction marks the beginning of the LTR after completion of reverse transcription) and also to avoid multiple discontinuities on the plus strand which then have to be repaired before integration. Added to this, the difference seen in priming in the presence of NC in vitro makes NC a likely candidate

for increasing the specificity of plus strand priming. Cell culture studies analyzing simultaneous mutations in both the PPT and NC have yet to be done. These could throw light on the relation between NC and plus strand priming during actual viral replication. For example, it would be of interest to see if a poorly binding NC mutant causes an increase in non specific priming during plus strand synthesis, or if PPT is replaced by a poorer primer can wild type NC cause a decrease in overall replication due to a decrease in plus strand priming efficiency.

The possibility that results seen in vitro do not have any biological significance cannot be ruled out. It is also possible that while the unwinding and competitive blocking of RT by NC are real properties of NC, the effect seen in vitro may not be necessary or have any significance in the host cell. Also, conditions used on the bench can never fully mimic what occurs in the cell and so the preferential inhibition of primers by NC could be due to the absence of some cellular condition or factor. This being said, it is important to note that most of the properties of NC were first studied in vitro before being confirmed in cells. Hence the relevance of NC's role in plus strand priming cannot be discarded.

The second part of this dissertation (Chapter 3) deals with the study of recombination under different NC and Mg^{2+} conditions. Recombination is a hallmark of HIV replication and its role in the evolution of the virus cannot be overstated. We were interested in studying the changes that occur in the pattern and extent of recombination as a result of variation in host (Mg^{2+}) and viral (NC) factors in vitro. The correlation of results obtained in vitro and those obtained from cell culture can help us identify conditions under which actual strand transfer occurs. This would be

potentially beneficial in developing more reliable in vitro assays to make it easier to study recombination and mechanisms underlying it.

The importance of NC to recombination, especially the first obligatory strand jump (30, 99) and its properties as a nucleic acid chaperone made its study with regard to location and pattern of strand transfer significant. Mg^{2+} was chosen as the other variable due to its effect on RT synthesis and RNA structure. Notably, the actual concentration of free Mg^{2+} during reverse transcription is unknown, which made the correlation of its effect on strand transfer in vitro and that observed in cells important. As seen from the results (Chapter 3), both NC and Mg^{2+} affect the pattern of recombination in the in vitro system used by us. Of the two, the effect of NC is more pronounced with regard to pausing and the location at which transfer occurs on the template while Mg^{2+} affects the extent of pausing and the efficiency of strand transfer. As expected, NC caused transfers to occur at earlier locations on the region of homology. It would be of interest to look at more variation in conditions, for example, increasing the concentration of NC in the reaction to see if the shift seen in transfers with increase in NC gets saturated at a higher level of NC. However, very high concentrations of NC in the reaction ($> 8 \mu M$) can cause inhibition of extension by RT (8).

Unfortunately, establishing the correlation between the in vitro and cell culture results was challenging. As described in the chapter discussion, the low number of sequences analyzed and the inherent differences such as template design etc. could be some reasons for the disparity seen. While aspects of the assay such as sample number and sequence design can possibly be optimized for better association,

it is possible that an in vitro experiment that efficiently mimics the cellular process may be unlikely. However, the advantages of using an effective in vitro system make its development crucial. In this regard, other parameters such as nucleotide levels, template RNA concentration etc. can also be examined. The fact that most cellular processes of HIV have first been studied in vitro underlines the importance of in vitro analyses. Another approach to the problem of correlation is to carry out in vitro analysis of a sequence that has already been studied in cell culture. One such example would be the acceptor-donor pair derived from the *env* region of the HIV genome. Cell culture studies have identified a recombination hotspot in the gp120 region of *env* (56). This pattern could be confirmed in vitro and analyzed under varying conditions of NC, Mg²⁺ etc.

While the study of recombination mechanisms and parameters would definitely help in the understanding of the process, the exact location of hotspots may not be very significant when it comes to actual recombination occurring in the virus. In the host, during recombination between subtypes or even variant members of the same origin, it is more likely that entire regions of the genome may be “hot zones” as opposed to single nucleotide hotspots. The location of hotspots is primarily dependent on the existence of strong structure and the subsequent pausing of RT. However, given the extremely condensed atmosphere of the core where replication and strand transfer occurs, it is likely that these structures may be opened up due to the presence of NC or that recombination occurs at regions of homology that are brought together by the aggregation properties of NC. It has also been shown that very high rates of recombination were localized in a region of almost no structure in vitro (36). This

further supports the idea that stretches of homology can also be powerful drivers of recombination in the absence of structure.

The significance of any recombination study would be to gain further insight into the evolutionary future of the virus. It would help us assess what inter-subtype recombinants are more likely to appear in a population than others and to develop prophylactic and therapeutic measures against imminent variants. The study of entire regions of higher strand transfer activity would be of more import in such cases where the entire genome is available for crossovers. Hence, it would be helpful to develop systems that make the analysis of larger stretches of RNA possible both in cells and the test tube.

Bibliography

1. March 15, 2007 2007, posting date. AVERT. [Online.]
2. 2008. International AIDS Vaccine Initiative.
3. June 6, 2006 2007, posting date. World Health Organization. [Online.]
4. **Abbondanzieri, E. A., G. Bokinsky, J. W. Rausch, J. X. Zhang, S. F. Le Grice, and X. Zhuang.** 2008. Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature* **453**:184-9.
5. **Allain, B., M. Lapadat-Tapolsky, C. Berlioz, and J. L. Darlix.** 1994. Transactivation of the minus-strand DNA transfer by nucleocapsid protein during reverse transcription of the retroviral genome. *Embo J* **13**:973-81.
6. **Amarasinghe, G. K., R. N. De Guzman, R. B. Turner, K. J. Chancellor, Z. R. Wu, and M. F. Summers.** 2000. 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* **301**:491-511.
7. **Amarasinghe, G. K., R. N. De Guzman, R. B. Turner, and M. F. Summers.** 2000. NMR structure of stem-loop SL2 of the HIV-1 psi RNA packaging signal reveals a novel A-U-A base-triple platform. *J Mol Biol* **299**:145-56.
8. **Anthony, R. M., and J. J. Destefano.** 2007. In vitro synthesis of long DNA products in reactions with HIV-RT and nucleocapsid protein. *J Mol Biol* **365**:310-24.
9. **Baba, S., K. Takahashi, Y. Koyanagi, N. Yamamoto, H. Takaku, R. J. Gorelick, and G. Kawai.** 2003. Role of the zinc fingers of HIV-1 nucleocapsid protein in maturation of genomic RNA. *J Biochem (Tokyo)* **134**:637-9.
10. **Baird, H. A., R. Galetto, Y. Gao, E. Simon-Loriere, M. Abreha, J. Archer, J. Fan, D. L. Robertson, E. J. Arts, and M. Negroni.** 2006. Sequence determinants of breakpoint location during HIV-1 intersubtype recombination. *Nucleic Acids Res* **34**:5203-16.
11. **Baltimore, D.** 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**:1209-11.
12. **Barin, F., M. F. McLane, J. S. Allan, T. H. Lee, J. E. Groopman, and M. Essex.** 1985. Virus envelope protein of HTLV-III represents major target antigen for antibodies in AIDS patients. *Science* **228**:1094-6.
13. **Barre-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Daguuet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**:868-71.

14. **Basu, V. P., M. Song, L. Gao, S. T. Rigby, M. N. Hanson, and R. A. Bambara.** 2008. Strand transfer events during HIV-1 reverse transcription. *Virus Res* **134**:19-38.
15. **Bebenek, K., J. Abbotts, S. H. Wilson, and T. A. Kunkel.** 1993. Error-prone polymerization by HIV-1 reverse transcriptase. *Journal of Biological Chemistry* **268**:10324-10334.
16. **Beltz, H., C. Clauss, E. Piemont, D. Ficheux, R. J. Gorelick, B. Roques, C. Gabus, J. L. Darlix, H. de Rocquigny, and Y. Mely.** 2005. Structural determinants of HIV-1 nucleocapsid protein for cTAR DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis. *J Mol Biol* **348**:1113-26.
17. **Brule, F., R. Marquet, L. Rong, M. A. Wainberg, B. P. Roques, S. F. Le Grice, B. Ehresmann, and C. Ehresmann.** 2002. 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* **8**:8-15.
18. **Buckman, J. S., W. J. Bosche, and R. J. Gorelick.** 2003. 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* **77**:1469-80.
19. **Carteau, S., S. C. Batson, L. Poljak, J. F. Mouscadet, H. de Rocquigny, J. L. Darlix, B. P. Roques, E. Kas, and C. Auclair.** 1997. Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg²⁺-dependent DNA integration in vitro. *J Virol* **71**:6225-9.
20. **Carteau, S., R. J. Gorelick, and F. D. Bushman.** 1999. Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. *J Virol* **73**:6670-9.
21. **Cen, S., Y. Huang, A. Khorchid, J. L. Darlix, M. A. Wainberg, and L. Kleiman.** 1999. The role of Pr55(gag) in the annealing of tRNA₃Lys to human immunodeficiency virus type 1 genomic RNA. *J Virol* **73**:4485-8.
22. **Charneau, P., M. Alizon, and F. Clavel.** 1992. A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *J Virol* **66**:2814-20.
23. **Charneau, P., and F. Clavel.** 1991. A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. *J Virol* **65**:2415-21.
24. **Chen, Y., M. Balakrishnan, B. P. Roques, and R. A. Bambara.** 2003. Steps of the acceptor invasion mechanism for HIV-1 minus strand strong stop transfer. *J Biol Chem* **278**:38368-75.
25. **Chin, M. P., T. D. Rhodes, J. Chen, W. Fu, and W. S. Hu.** 2005. Identification of a major restriction in HIV-1 intersubtype recombination. *Proc Natl Acad Sci U S A* **102**:9002-7.
26. **Coffin, J. M., S. H. Hughes, and H. E. Varmus.** 1997. *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27. **Craigie, R.** 2001. HIV integrase, a brief overview from chemistry to therapeutics. *J Biol Chem* **276**:23213-6.

28. **Cristofaro, J. V., J. W. Rausch, S. F. Le Grice, and J. J. DeStefano.** 2002. Mutations in the ribonuclease H active site of HIV-RT reveal a role for this site in stabilizing enzyme-primer-template binding. *Biochemistry* **41**:10968-10975.
29. **D'Orso, I., J. R. Grunwell, R. L. Nakamura, C. Das, and A. D. Frankel.** 2008. Targeting tat inhibitors in the assembly of human immunodeficiency virus type 1 transcription complexes. *J Virol* **82**:9492-504.
30. **Darlix, J. L., A. Vincent, C. Gabus, H. de Rocquigny, and B. Roques.** 1993. Trans-activation of the 5' to 3' viral DNA strand transfer by nucleocapsid protein during reverse transcription of HIV1 RNA. *C R Acad. Sci. III* **316**:763-71.
31. **De Guzman, R. N., R. B. Turner, and M. F. Summers.** 1998. Protein-RNA recognition. *Biopolymers* **48**:181-95.
32. **De Guzman, R. N., Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, and M. F. Summers.** 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **279**:384-8.
33. **de Soultrait, V. R., A. Caumont, V. Parissi, N. Morellet, M. Ventura, C. Lenoir, S. Litvak, M. Fournier, and B. Roques.** 2002. A novel short peptide is a specific inhibitor of the human immunodeficiency virus type 1 integrase. *J Mol Biol* **318**:45-58.
34. **de Soultrait, V. R., P. Y. Lozach, R. Altmeyer, L. Tarrago-Litvak, S. Litvak, and M. L. Andreola.** 2002. DNA aptamers derived from HIV-1 RNase H inhibitors are strong anti-integrase agents. *J Mol Biol* **324**:195-203.
35. **Demirov, D. G., and E. O. Freed.** 2004. Retrovirus budding. *Virus Res* **106**:87-102.
36. **Derebail, S. S., and J. J. Destefano.** 2004. Mechanistic Analysis of Pause Site-dependent and -independent Recombinogenic Strand Transfer from Structurally Diverse Regions of the HIV Genome. *J. Biol. Chem.* **279**:47446-54.
37. **DeStefano, J. J.** 1995. Human immunodeficiency virus nucleocapsid protein stimulates strand transfer from internal regions of heteropolymeric RNA templates. *Arch Virol* **140**:1775-89.
38. **DeStefano, J. J.** 1996. 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.* **271**:16350-16356.
39. **DeStefano, J. J.** 1995. The orientation of binding of human immunodeficiency virus reverse transcriptase on nucleic acid hybrids. *Nucleic Acids Res* **23**:3901-8.
40. **DeStefano, J. J., R. A. Bambara, and P. J. Fay.** 1993. Parameters that influence the binding of human immunodeficiency virus reverse transcriptase to nucleic acid structures. *Biochemistry* **32**:6908-6915.
41. **DeStefano, J. J., and J. V. Cristofaro.** 2006. Selection of primer-template sequences that bind human immunodeficiency virus reverse transcriptase with high affinity. *Nucleic Acids Res.* **34**:130-9.

42. **DeStefano, J. J., J. V. Cristofaro, S. Derebail, W. P. Bohlayer, and M. J. Fitzgerald-Heath.** 2001. Physical mapping of HIV reverse transcriptase to the 5' end of RNA primers. *J Biol Chem* **276**:32515-21.
43. **DeStefano, J. J., L. M. Mallaber, P. J. Fay, and R. A. Bambara.** 1993. Determinants of the RNase H cleavage specificity of human immunodeficiency virus reverse transcriptase. *Nucleic Acids Res* **21**:4330-4338.
44. **DeStefano, J. J., L. M. Mallaber, P. J. Fay, and R. A. Bambara.** 1994. Quantitative analysis of RNA cleavage during RNA-directed DNA synthesis by human immunodeficiency and avian myeloblastosis virus reverse transcriptases. *Nucleic Acids Res* **22**:3793-800.
45. **DeStefano, J. J., L. M. Mallaber, L. Rodriguez-Rodriguez, P. J. Fay, and R. A. Bambara.** 1992. Requirements for strand transfer between internal regions of heteropolymer templates by human immunodeficiency virus reverse transcriptase. *J Virol* **66**:6370-8.
46. **di Marzo Veronese, F., T. D. Copeland, A. L. DeVico, R. Rahman, S. Oroszlan, R. C. Gallo, and M. G. Sarngadharan.** 1986. Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. *Science* **231**:1289-91.
47. **Dirac, A. M., H. Huthoff, J. Kjems, and B. Berkhout.** 2002. Requirements for RNA heterodimerization of the human immunodeficiency virus type 1 (HIV-1) and HIV-2 genomes. *J Gen Virol* **83**:2533-42.
48. **Drummond, J. E., P. Mounts, R. J. Gorelick, J. R. Casas-Finet, W. J. Bosche, L. E. Henderson, D. J. Waters, and L. O. Arthur.** 1997. Wild-type and mutant HIV type 1 nucleocapsid proteins increase the proportion of long cDNA transcripts by viral reverse transcriptase. *AIDS Res Hum Retroviruses* **13**:533-43.
49. **Egele, C., E. Piemont, P. Didier, D. Ficheux, B. Roques, J. L. Darlix, H. de Rocquigny, and Y. Mely.** 2007. The single-finger nucleocapsid protein of moloney murine leukemia virus binds and destabilizes the TAR sequences of HIV-1 but does not promote efficiently their annealing. *Biochemistry* **46**:14650-62.
50. **FDA.** 2008. Drugs Used in the Treatment of HIV Infection.
51. **Feng, Y. X., S. Campbell, D. Harvin, B. Ehresmann, C. Ehresmann, and A. Rein.** 1999. 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* **73**:4251-6.
52. **Feng, Y. X., T. D. Copeland, L. E. Henderson, R. J. Gorelick, W. J. Bosche, J. G. Levin, and A. Rein.** 1996. HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro. *Proc Natl Acad Sci U S A* **93**:7577-81.
53. **Fischl, M. A., D. D. Richman, M. H. Grieco, M. S. Gottlieb, P. A. Volberding, O. L. Laskin, J. M. Leedom, J. E. Groopman, D. Mildvan, R. T. Schooley, and et al.** 1987. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. *N Engl J Med* **317**:185-91.

54. **Fuentes, G. M., L. Rodriguez-Rodriguez, P. J. Fay, and R. A. Bambara.** 1995. Use of an oligoribonucleotide containing the polypurine tract sequence as a primer by HIV reverse transcriptase. *J Biol Chem* **270**:28169-76.
55. **Galetto, R., V. Giacomoni, M. Veron, and M. Negroni.** 2006. Dissection of a circumscribed recombination hot spot in HIV-1 after a single infectious cycle. *J Biol Chem* **281**:2711-20.
56. **Galetto, R., A. Moumen, V. Giacomoni, M. Veron, P. Charneau, and M. Negroni.** 2004. The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo. *J Biol Chem* **279**:36625-32.
57. **Gallo, R. C., S. Z. Salahuddin, M. Popovic, G. M. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Redfield, J. Oleske, B. Safai, and et al.** 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**:500-3.
58. **Goff, S. P.** 1990. Retroviral reverse transcriptase: synthesis, structure and function. *Acquired Immune Deficiency Syndromes* **3**:817-831.
59. **Goldschmidt, V., J. Didierjean, B. Ehresmann, C. Ehresmann, C. Isel, and R. Marquet.** 2006. Mg²⁺ dependency of HIV-1 reverse transcription, inhibition by nucleoside analogues and resistance. *Nucleic Acids Res.* **34**:42-52.
60. **Gorelick, R. J., D. J. Chabot, A. Rein, L. E. Henderson, and L. O. Arthur.** 1993. The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. *J Virol* **67**:4027-36.
61. **Grossman, Z., M. Meier-Schellersheim, W. E. Paul, and L. J. Picker.** 2006. Pathogenesis of HIV infection: what the virus spares is as important as what it destroys. *Nat Med* **12**:289-95.
62. **Guo, J., L. E. Henderson, J. Bess, B. Kane, and J. G. Levin.** 1997. 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* **71**:5178-88.
63. **Guo, J., T. Wu, J. Anderson, B. F. Kane, D. G. Johnson, R. J. Gorelick, L. E. Henderson, and J. G. Levin.** 2000. Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J Virol* **74**:8980-8.
64. **Guo, J., T. Wu, B. F. Kane, D. G. Johnson, L. E. Henderson, R. J. Gorelick, and J. G. Levin.** 2002. 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* **76**:4370-8.
65. **Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon.** 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature* **326**:662-9.
66. **Heath, M. J., S. S. Derebail, R. J. Gorelick, and J. J. DeStefano.** 2003. Differing roles of the N-terminal and C-terminal zinc fingers in HIV-1 nucleocapsid protein enhanced nucleic acid annealing. *J. Biol. Chem.* **278**:30755-30763.

67. **Hou, E. W., R. Prasad, W. A. Beard, and S. H. Wilson.** 2004. High-level expression and purification of untagged and histidine-tagged HIV-1 reverse transcriptase. *Protein Expr Purif* **34**:75-86.
68. **Hsiou, Y., J. Ding, K. Das, A. D. Clark, Jr., S. H. Hughes, and E. Arnold.** 1996. Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure* **4**:853-60.
69. **Hu, W. S., T. Rhodes, Q. Dang, and V. Pathak.** 2003. Retroviral recombination: review of genetic analyses. *Front Biosci* **8**:d143-55.
70. **Hu, W. S., and H. M. Temin.** 1990. Retroviral recombination and reverse transcription. *Science* **250**:1227-33.
71. **Huet, T., R. Cheynier, A. Meyerhans, G. Roelants, and S. Wain-Hobson.** 1990. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345**:356-9.
72. **Hwang, C. K., E. S. Svarovskaia, and V. K. Pathak.** 2001. Dynamic copy choice: steady state between murine leukemia virus polymerase and polymerase-dependent RNase H activity determines frequency of in vivo template switching. *Proc Natl Acad Sci U S A* **98**:12209-14.
73. **Jacob, D. T., and J. J. DeStefano.** 2008. A new role for HIV nucleocapsid protein in modulating the specificity of plus strand priming. *Virology* **378**:385-96.
74. **Jacobo-Molina, A., J. Ding, R. G. Nanni, A. D. Clark, Jr., X. Lu, C. Tantillo, R. L. Williams, G. Kamer, A. L. Ferris, P. Clark, and et al.** 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc Natl Acad Sci U S A* **90**:6320-4.
75. **Ji, X., G. J. Klarmann, and B. D. Preston.** 1996. Effect of human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein on HIV-1 reverse transcriptase activity in vitro. *Biochemistry* **35**:132-143.
76. **Johnson, P. E., R. B. Turner, Z. R. Wu, L. Hairston, J. Guo, J. G. Levin, and M. F. Summers.** 2000. A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry* **39**:9084-91.
77. **Khan, R., and D. P. Giedroc.** 1992. Recombinant human immunodeficiency virus type 1 nucleocapsid (NCp7) protein unwinds tRNA. *J Biol Chem* **267**:6689-95.
78. **Klarmann, G. J., H. Yu, X. Chen, J. P. Dougherty, and B. D. Preston.** 1997. Discontinuous plus-strand DNA synthesis in human immunodeficiency virus type 1-infected cells and in a partially reconstituted cell-free system. *J. Virol.* **71**:9259-9269.
79. **Kung, H. J., Y. K. Fung, J. E. Majors, J. M. Bishop, and H. E. Varmus.** 1981. Synthesis of plus strands of retroviral DNA in cells infected with avian sarcoma virus and mouse mammary tumor virus. *J Virol* **37**:127-38.
80. **Kunkel, T. A., and S. H. Wilson.** 1998. DNA polymerases on the move. *Nat Struct Biol* **5**:95-9.

81. **Kvaratskhelia, M., S. R. Budihas, and S. F. Le Grice.** 2002. Pre-existing distortions in nucleic acid structure aid polypurine tract selection by HIV-1 reverse transcriptase. *J Biol Chem* **277**:16689-96.
82. **Lambele, M., B. Labrosse, E. Roch, A. Moreau, B. Verrier, F. Barin, P. Roingeard, F. Mammano, and D. Brand.** 2007. Impact of natural polymorphism within the gp41 cytoplasmic tail of human immunodeficiency virus type 1 on the intracellular distribution of envelope glycoproteins and viral assembly. *J Virol* **81**:125-40.
83. **Lapadat-Tapolsky, M., H. De Rocquigny, D. Van Gent, B. Roques, R. Plasterk, and J. L. Darlix.** 1993. Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle [published erratum appears in *Nucleic Acids Res* 1993 Apr 25;21(8):2024]. *Nucleic Acids Res* **21**:831-9.
84. **Laughrea, M., N. Shen, L. Jette, J. L. Darlix, L. Kleiman, and M. A. Wainberg.** 2001. Role of distal zinc finger of nucleocapsid protein in genomic RNA dimerization of human immunodeficiency virus type 1; no role for the palindrome crowning the R-U5 hairpin. *Virology* **281**:109-16.
85. **Le Cam, E., D. Coulaud, E. Delain, P. Petitjean, B. P. Roques, D. Gerard, E. Stoylova, C. Vuilleumier, S. P. Stoylov, and Y. Mely.** 1998. Properties and growth mechanism of the ordered aggregation of a model RNA by the HIV-1 nucleocapsid protein: an electron microscopy investigation. *Biopolymers* **45**:217-29.
86. **Le Novere, N.** 2001. MELTING, computing the melting temperature of nucleic acid duplex. *Bioinformatics* **17**:1226-7.
87. **Lener, D., V. Tanchou, B. P. Roques, S. F. Le Grice, and J. L. Darlix.** 1998. Involvement of HIV-1 nucleocapsid protein in the recruitment of reverse transcriptase into nucleoprotein complexes formed in vitro. *J Biol Chem* **273**:33781-6.
88. **Levin, J. G., J. Guo, I. Rouzina, and K. Musier-Forsyth.** 2005. Nucleic Acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acids Res. Mol. Biol.* **80**:217-286.
89. **Levy, D. N., G. M. Aldrovandi, O. Kutsch, and G. M. Shaw.** 2004. Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci U S A* **101**:4204-9.
90. **Levy, J. A., A. D. Hoffman, S. M. Kramer, J. A. Landis, J. M. Shimabukuro, and L. S. Oshiro.** 1984. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* **225**:840-2.
91. **Miles, L. R., B. E. Agresta, M. B. Khan, S. Tang, J. G. Levin, and M. D. Powell.** 2005. Effect of polypurine tract (PPT) mutations on human immunodeficiency virus type 1 replication: a virus with a completely randomized PPT retains low infectivity. *J Virol* **79**:6859-67.
92. **Miller, M. D., B. Wang, and F. D. Bushman.** 1995. Human immunodeficiency virus type 1 preintegration complexes containing discontinuous plus strands are competent to integrate in vitro. *J Virol* **69**:3938-44.

93. **Myers, G., K. MacInnes, and B. Korber.** 1992. The emergence of simian/human immunodeficiency viruses. *AIDS Res Hum Retroviruses* **8**:373-86.
94. **Narayanan, N., R. J. Gorelick, and J. J. DeStefano.** 2006. Structure/function mapping of amino acids in the N-terminal zinc finger of the human immunodeficiency virus type 1 nucleocapsid protein: residues responsible for nucleic acid helix destabilizing activity. *Biochemistry* **45**:12617-28.
95. **Negroni, M., and H. Buc.** 2001. Retroviral recombination: what drives the switch? *Nat. Rev. Mol. Cell Biol.* **2**:151-5.
96. **Onafuwa, A., W. An, N. D. Robson, and A. Telesnitsky.** 2003. Human Immunodeficiency Virus Type 1 Genetic Recombination Is More Frequent Than That of Moloney Murine Leukemia Virus despite Similar Template Switching Rates. *J. Virol.* **77**:4577-4587.
97. **Palaniappan, C., J. K. Kim, M. Wisniewski, P. J. Fay, and R. A. Bambara.** 1998. Control of initiation of viral plus strand DNA synthesis by HIV reverse transcriptase. *J Biol Chem* **273**:3808-16.
98. **Pantaleo, G., J. F. Demarest, T. Schacker, M. Vaccarezza, O. J. Cohen, M. Daucher, C. Graziosi, S. S. Schnittman, T. C. Quinn, G. M. Shaw, L. Perrin, G. Tambussi, A. Lazzarin, R. P. Sekaly, H. Soudeyns, L. Corey, and A. S. Fauci.** 1997. The qualitative nature of the primary immune response to HIV infection is a prognosticator of disease progression independent of the initial level of plasma viremia. *Proc Natl Acad Sci U S A* **94**:254-8.
99. **Peliska, J. A., S. Balasubramanian, D. P. Giedroc, and S. J. Benkovic.** 1994. Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity. *Biochemistry* **33**:13817-23.
100. **Pollard, V. W., and M. H. Malim.** 1998. The HIV-1 Rev protein. *Annu Rev Microbiol* **52**:491-532.
101. **Powell, M. D., and J. G. Levin.** 1996. Sequence and structural determinants required for priming of plus-strand DNA synthesis by the human immunodeficiency virus type 1 polypurine tract. *J Virol* **70**:5288-96.
102. **Raja, A., and J. J. DeStefano.** 1999. Kinetic analysis of the effect of HIV nucleocapsid protein (NCp) on internal strand transfer reactions. *Biochemistry* **38**:5178-84.
103. **Ramboarina, S., S. Druillenec, N. Morellet, S. Bouaziz, and B. P. Roques.** 2004. Target specificity of human immunodeficiency virus type 1 NCp7 requires an intact conformation of its CCHC N-terminal zinc finger. *J Virol* **78**:6682-7.
104. **Ramirez, B. C., E. Simon-Loriere, R. Galetto, and M. Negroni.** 2008. Implications of recombination for HIV diversity. *Virus Res* **134**:64-73.
105. **Rausch, J. W., and S. F. Le Grice.** 2004. 'Binding, bending and bonding': polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. *Int. J. Biochem. Cell Biol.* **36**:1752-66.

106. **Rein, A., D. P. Harvin, J. Mirro, S. M. Ernst, and R. J. Gorelick.** 1994. Evidence that a central domain of nucleocapsid protein is required for RNA packaging in murine leukemia virus. *J Virol* **68**:6124-9.
107. **Robertson, D. L., B. H. Hahn, and P. M. Sharp.** 1995. Recombination in AIDS viruses. *J Mol Evol* **40**:249-59.
108. **Rodriguez-Rodriguez, L., Z. Tsuchihashi, G. M. Fuentes, R. A. Bambara, and P. J. Fay.** 1995. Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro. *J Biol Chem* **270**:15005-11.
109. **Rong, L., C. Liang, M. Hsu, L. Kleiman, P. Petitjean, H. de Rocquigny, B. P. Roques, and M. A. Wainberg.** 1998. 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* **72**:9353-8.
110. **Sambrook, J., and D. W. Russell.** 2001. *Molecular Cloning: A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
111. **Sarafianos, S. G., K. Das, C. Tantillo, A. D. Clark, Jr., J. Ding, J. M. Whitcomb, P. L. Boyer, S. H. Hughes, and E. Arnold.** 2001. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *Embo J* **20**:1449-61.
112. **Schatz, O., F. V. Cromme, F. Gruninger-Leitch, and S. F. J. Le Grice.** 1989. Point mutations in conserved amino acid residues within the C-terminal domain of HIV-1 reverse transcriptase specifically repress RNase H function. *FEBS Letters* **257**:311-314.
113. **Schultz, S. J., M. Zhang, and J. J. Champoux.** 2006. Sequence, distance, and accessibility are determinants of 5'-end-directed cleavages by retroviral RNases H. *J Biol Chem* **281**:1943-55.
114. **Schultz, S. J., M. Zhang, C. D. Kelleher, and J. J. Champoux.** 2000. Analysis of plus-strand primer selection, removal, and reutilization by retroviral reverse transcriptases. *J Biol Chem* **275**:32299-309.
115. **Shehu-Xhilaga, M., S. M. Crowe, and J. Mak.** 2001. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J Virol* **75**:1834-41.
116. **Shehu-Xhilaga, M., H. G. Kraeusslich, S. Pettit, R. Swanstrom, J. Y. Lee, J. A. Marshall, S. M. Crowe, and J. Mak.** 2001. Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. *J Virol* **75**:9156-64.
117. **Starnes, M. C., and Y. C. Cheng.** 1989. Human immunodeficiency virus reverse transcriptase-associated RNase H activity. *J Biol Chem* **264**:7073-7.
118. **Stoylov, S. P., C. Vuilleumier, E. Stoylova, H. De Rocquigny, B. P. Roques, D. Gerard, and Y. Mely.** 1997. Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. *Biopolymers* **41**:301-12.
119. **Tanchou, V., D. Decimo, C. Pechoux, D. Lener, V. Rogemond, L. Berthoux, M. Ottmann, and J. L. Darlix.** 1998. Role of the N-terminal zinc

- finger of human immunodeficiency virus type 1 nucleocapsid protein in virus structure and replication. *J Virol* **72**:4442-7.
120. **Tanchou, V., T. Delaunay, M. Bodeus, B. Roques, J. L. Darlix, and R. Benarous.** 1995. Conformational changes between human immunodeficiency virus type 1 nucleocapsid protein NCp7 and its precursor NCp15 as detected by anti- NCp7 monoclonal antibodies. *J Gen Virol* **76**:2457-66.
 121. **Tanchou, V., C. Gabus, V. Rogemond, and J.-L. Darlix.** 1995. Formation of stable and functional HIV-1 nucleoprotein complexes *in vitro*. *J. Mol. Biol.* **252**:563-571.
 122. **Temin, H. M., and S. Mizutani.** 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**:1211-3.
 123. **ter Brake, O., P. Konstantinova, M. Ceylan, and B. Berkhout.** 2006. Silencing of HIV-1 with RNA interference: a multiple shRNA approach. *Mol Ther* **14**:883-92.
 124. **Williams, M. C., R. J. Gorelick, and K. Musier-Forsyth.** 2002. Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proc. Natl. Acad. Sci. U. S. A.* **99**:8614-9.
 125. **Wisniewski, M., M. Balakrishnan, C. Palaniappan, P. J. Fay, and R. A. Bambara.** 2000. Unique progressive cleavage mechanism of HIV reverse transcriptase RNase H. *Proc. Natl. Acad. Sci. U S A* **97**:11978-11983.
 126. **Wisniewski, M., Y. Chen, M. Balakrishnan, C. Palaniappan, B. P. Roques, P. J. Fay, and R. A. Bambara.** 2002. Substrate requirements for secondary cleavage by HIV-1 reverse transcriptase RNase H. *J Biol Chem* **277**:28400-10.
 127. **Wu, T., S. L. Heilman-Miller, and J. G. Levin.** 2007. Effects of nucleic acid local structure and magnesium ions on minus-strand transfer mediated by the nucleic acid chaperone activity of HIV-1 nucleocapsid protein. *Nucleic Acids Res* **35**:3974-87.
 128. **You, J. C., and C. S. McHenry.** 1993. HIV nucleocapsid protein. Expression in *Escherichia coli*, purification, and characterization. *J Biol Chem* **268**:16519-27.
 129. **You, J. C., and C. S. McHenry.** 1994. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J Biol Chem* **269**:31491-5.
 130. **Zhu, T., B. T. Korber, A. J. Nahmias, E. Hooper, P. M. Sharp, and D. D. Ho.** 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**:594-7.
 131. **Zhuang, J., A. E. Jetzt, G. Sun, H. Yu, G. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty.** 2002. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol* **76**:11273-82.
 132. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**:3406-15.