ABSTRACT

Title of Dissertation: INFLUENCE OF VIRAL NUCLEOCAPSID

PROTEIN   AND   GENOMIC   RNA

STRUCTURAL   INTRICACIES   ON   THE

MECHANISM OF HIV RECOMBINATION

Suchitra S Derebail, Doctor of Philosophy, 2004

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

Internal strand transfers involve template switching during retroviral

replication, within internal regions of the viral genome. Such transfers are a major

source of genetic variability in retroviruses like HIV-1. An in vitro strand transfer

assay that mimicked recombinational events occurring during reverse transcription in

HIV-1 was used to assess the role of nucleocapsid protein (NC) and structural

intricacies of genomic RNA in strand transfer.   Transfers in highly structured

templates from the U3 3' LTR, gag-pol frameshift region, and Rev response element

(RRE) were strongly enhanced by NC. In contrast, weakly structured templates from

the env and pol-vif regions transferred well without NC and showed lower

enhancement.   Assays conducted using NC zinc finger mutants supported a

differential role for the two fingers in strand transfer with finger one (N-terminal) being more important on highly structured RNAs. The lack of strong polymerase pause sites in the weakly structured templates from the env and pol-vif regions demonstrated that non-pause driven mechanisms could also promote transfer.

Mapping assays were conducted on high and low structured templates from the gag-pol and the env regions respectively (called GagPol and Env templates), to locate the point(s) of transfer in each case. The majority of transfers were located near a major pause site in the gag-pol region; in contrast, in the env region, most transfers were located towards the end of the homologous region between donor and acceptor templates. Various truncated/mutant GagPol acceptor templates were analyzed with wild type GagPol donor templates in strand transfer assays. Results indicated that destabilized acceptor templates enhance the level of transfer and cause a highly efficient 'chasing' of the pause site into transfer products. The outcome of these experiments also suggested that strand transfer in the GagPol templates is via a pause induced 'donor dissociation' method. In the Env template the mechanism of transfer was proposed to occur by a pause independent, 'acceptor invasion' method. This knowledge about the interplay of RNA structure and NC protein on recombination could help in designing antiviral vaccines and drug inhibitors.

INFLUENCE OF VIRAL NUCLEOCAPSID PROTEIN AND GENOMIC RNA

STRUCTURAL INTRICACIES ON THE MECHANISM OF HIV

RECOMBINATION

by

Suchitra S. Derebail

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
2004

Advisory Committee:

     Dr. Jeffrey J. DeStefano, Chairman/Advisor
     Dr. Barbara Thorne
     Dr. Eric Baehrecke
     Dr. James Culver
     Dr. Siba K. Samal

DEDICATION

This thesis is dedicated to my Pappa and Amma (Sudhakar and Ramaa) and to my

dear husband Pameet for all their love and support.

ACKNOWLEDGEMENTS

When I look back at the end of this long journey, I am extremely grateful for all those people without whom this journey would have been longer! First of all, I want to thank my advisor Dr Jeffrey DeStefano. When I came into your lab, I couldn't tell the difference between a pipetman and a screwdriver. Thanks to your training and guidance, now I have good hands for molecular biology. Over the years, I have also come to appreciate the way you have achieved a good balance between academics and sports in your life. I bet no one can leave your lab without being bitten by the 'running' bug. I run almost everyday and Megan has already run the marathon. I hope this tradition continues in the lab.

I owe huge thanks to the great banking institution called HongKong Bank in Chennai, India. Without my job at this bank I could have never dreamt of funding my ticket to go all the way across the globe. I was very lucky to have had friends like Sumithra, and Kiran. Thank you both for helping me through some rough times during my three year stint at the bank. Sumi, you are a wonderful human being, you always look for the best in others and I have never met anyone else like you. Kiran, thanks for convincing me that it was a not a crazy idea to throw away a good paying job to go back to school.

Thank you Jaya akka for receiving me at the airport. Thank you for giving me all the clothes, kitchenware and bed things. Without the comfort of knowing you were nearby, I couldn't have survived the first few months in this new country. I

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

HIV	Human immunodeficiency virus
AMV	Avian myeloblastosis virus
MMLV	Moloney murine leukemia virus
HTLV	Human T-cell leukemia virus
AIDS	Acquired immunodeficiency syndrome
LTR	Long terminal repeat
RT	Reverse transcriptase
RNaseH	Ribonuclease H
PBS	Primer binding site
PPT	Poly-purine tract
AMP	Adenosine mononucleoside phosphate
rNTPs	Ribonucleoside triphosphates
dNTPs	Deoxynucleotide triphosphates
DTT	Dithiothreitol
PNK	Polynucleotide kinase
EDTA	Ethylenediaminetetraacetic acid
CIP	Calf intestinal phosphatase
PCR	Polmerase chain reaction
ss	Single-stranded
ds	Double-stranded
IN	Integrase protein
PR	Protease protein

NC   Nucleocapsid protein

TAR   *trans*-activation response

RRE   Rev response element

Chapter 1   Acquired   Immunodeficiency   Virus;   History, Pathogenesis and Life cycle

1.1 Introduction

Human immunodeficiency virus (HIV) is a member of Lentiviruses that comprise a separate genus of the family Retroviridae (retroviruses). HIV (Human Immunodeficiency Virus) is the causative agent of one of the world's deadliest scourges - AIDS (Acquired Immunodeficiency Syndrome). AIDS is characterized by a severe impairment of the immune system and is the term given for any or all of some 27 known diseases and symptoms. A person is diagnosed with AIDS if he/she has any of those 27 diseases or a T4 lymphocyte count of less than 200/µl of blood and also tests positive for antibodies to HIV. The virus was first isolated by Luc Montagnier and his colleagues at the Pasteur Institute in Paris in January of 1983 and was named lymphadenopathy-associated virus (LAV) [1]. It was also isolated by Robert Gallo and named Human T Cell Lymphotropic Virus III (HTLV III) [2]. In later years, the term HIV was coined by the Committee on the Taxonomy of Viruses for worldwide use. HIV-1 and HIV-2 are the two types of the virus; the latter is found mostly in West Africa but HIV-1 is far more common. HIV-1 is categorized into 3 groups; Group M or Main comprises all the main genotypes or clades found in different populations and Group O or Outlier comprises all those that are different from Group M. Group N or New strain was found in people living in Cameroon of West Africa. The main group is further classified in 10 subtypes or clades: A-J. All of the above mentioned types are enveloped, plus stranded RNA retroviruses so named because of the reverse or 'retro' flow of genetic information from RNA to

DNA during their replication cycle. It is important to note that all the work done in this report bears relevance to the HIV-1 virus although the similarities between the two viruses suggest that the general conclusions would hold for both.

1.2   AIDS Epidemiology

Since the first clinical evidence of HIV/AIDS was reported in 1981, the AIDS pandemic has claimed nearly 20 million people worldwide [3]. Currently, there are an estimated number of 40 million people globally that are living with the virus. In the year 2003 alone, approximately 3.1 million lives succumbed to the diseases caused by AIDS and an estimated 5 million people newly acquired the virus, 800,000 of them children [3]. Enabling the uninfected individuals to protect themselves against HIV, and providing adequate and affordable treatment to people living with the virus, represent two of the biggest challenges facing humankind today.

1.3   Immunopathogenesis of HIV infection

Upon entry of the body *via* infected body fluids like blood, semen, and vaginal secretions, HIV infects T4 lymphocytes. The T4 cell count in uninfected adults and adolescents is approximately $1000/\mu l$ of blood. A normal T4 cell is crucial for activating other main classes of lymphocytes like B cells, natural killer cells etc. It does this by secreting lymphokines when activated by phagocytes. As HIV takes over the T4 cells, the resultant depletion triggers a profound immunosuppression. At this juncture, the host becomes susceptible to attacks from opportunistic infections, cancers/organ failures etc [4].

1.4 Opportunistic infections (OIs) and cancers in AIDS patients

OIs are the primary threat to AIDS patients. About 90% of AIDS related deaths are caused by OIs, 7% are due to cancer and only 3% due to other causes. A wide range of viruses, bacteria, fungi and protozoa that harmlessly inhabit a healthy individual become pathogenic in an AIDS patient due to collapse of the immune system. Some of the common AIDS related OIs are candidal esophagitis, PCP (Pnuemocystis carinii pneumonia) and cytomegalovirus infections. About 40% of AIDS patients develop cancers like Burkitt's lymphoma, squamous cell carcinoma and Kaposi's sarcoma (KS). Of these KS is the most frequent infection; it comes on quickly and spreads aggressively to the gut, lung, pleura and the lymph nodes [5].

1.5 Anti-HIV therapy

The three main classes of anti-HIV drugs are:

- Nucleoside analog RT inhibitors (NRTI) for example Zidovudine, Didanosine; these are nucleoside analogs that inhibit HIV RT enzyme that functions early in the HIV life cycle. They act by incorporating into the elongating strand of HIV DNA and causing chain termination.

- Non-nucleoside analog RT inhibitors (NNRTI) for example Nevrapine, Delavirdine; they also inhibit RT but unlike NRTIs they inhibit replication by binding noncompetitively to RT.

- Protease inhibitors for example Saquinavir, Ritonavir; these drugs work by inhibiting the HIV protease enzymes that function late in the HIV life cycle. The lack of proteases results in production of non-infectious virions.

All the above drugs are administered as multi-drug cocktails to combat the emergence of drug resistant mutants. This kind of combination therapy is referred to as HAART (Highly active antiretroviral therapy). To date there is only one FDA-approved vaccine in phase III trials; AIDSVAX [6]. This vaccine uses recombinant gp120 which is a glycoprotein on the envelope. But preliminary results of the vaccine (conducted recently in Nov 2003) have yielded no positive results [7]. Attenuated vaccines are being held back because there is no 100% guarantee that all viruses in the vaccine will be inactivated. Currently, there are 74 experimental vaccines in animal testing and basic research. Most of these target surface proteins of HIV. Therapeutic vaccines like Remune that aim to bring infection under long-term control are undergoing trials in US, Europe and Thailand [8].

1.6 Hallmarks of HIV life cycle

Genetic structure and composition of HIV virions: Generally, retroviral particles are composed of 1-2% RNA, 35% lipids and 65% protein [9]. In pure viral preparations, three quarters of the total protein component is due to internal structural proteins, the rest is composed of viral envelope proteins, viral enzymes and some cellular proteins. Two thirds of the RNA content is genomic in nature, the remainder is non-genomic RNA accounted for by nearly 100 copies of transfer RNAs, some 5S rRNA 7S RNA and traces of cellular mRNAs. The genomic RNA contains two completely identical or nearly identical copies of the single stranded, positive sense RNA genome that are held together near the 5' end by multiple regions of base pairing (Figure 1-1). Though two copies are encapsidated, in general only one provirus is detected in a single virion infection, hence retroviruses are viewed as

pseudodiploids [10]. The existence of two genomes probably helps the retroviruses survive damage and also accounts for the high rate of genetic recombination seen in these viruses [10]. The genomic RNA is coated along its entire length with a nucleocapsid protein (NC). This is a small basic protein that can bind nonspecifically to RNA and DNA with approximately one molecule for every seven nucleotides [11]. The HIV genome of 9,749 nucleotides contains nine genes which produce at least nine individual proteins (Figure 1-2). These proteins are classified into; three structural polyproteins encoded by *gag*, *env* and *pol* genes, two regulatory proteins encoded by tat and rev genes and four accessory proteins encoded by *nef*, *vif*, *vpu*, and *vpr*. The *gag* (group-specific antigen) gene that codes for internal structural proteins; MA (Matrix), CA (Capsid), and NC (Nucleocapsid) has the potential to direct viral particle formation even in the absence of the other genes. The *pol* genes code for the viral enzymes RT (Reverse Transcriptase), PR (Protease) and IN (Integrase), while the *env* gene codes for two surface glycoproteins; gp120 forms the external "spikes" of the virus and gp41 forms the transmembrane protein that connects gp120 to the viral capsid surface. Table 1-1 lists the different genes, their proteins and functions.

Viral entry: The envelope complex of most retroviruses is a knobbed spike that comprises two polypeptides; an external glycosylated surface polypeptide (gp120/SU) and a membrane-spanning protein (gp41/TM), both of which are encoded by the *env* gene (refer Figure 1-1). During viral entry, SU binds to a specific receptor molecule on the target cell which induces TM to fuse with the cell membrane. For HIV and SIV, the primary receptor molecule is the 60-kDa CD4 antigen found on T-helper cells, macrophages, monocytes and other phagocytic cells [12-14]. The work

of Maddon *et al.* [15] showed that transfection and expression of cDNA for human CD4 in HeLa cells rendered these resistant cells permissive for HIV infection. Initially, HIV was shown to enter cells by endocytosis, later on, it was shown that entry is *via* a pH-independent method [16]. In addition to CD4 receptor molecule, viral entry requires a second receptor called the β-chemokine receptor. These are of two types; the "FUSIN" or CXCKR4 receptor present on T-cells and the CCR5 found on macrophages. All isolates of HIV are classified into T-tropic or M-tropic isolates depending on the type of chemokine receptor they use for entry.

Table 1-1: Genes of HIV - 1

Gene name(s)	Molecular Mass (kDa)	Function
gag (group antigen)	p17 p24 p9, p6	Matrix (MA) for protection Capsid (CA) for protection Nucleocapsid (NC) for protection
pol (polymerase)	p10 p31 p50	Protease enzyme cleaves the viral polyproteins Integrase enzyme integrates proviral DNA into host cell DNA Reverse transcriptase enzyme converts viral genome into double stranded provirus
env (envelope)	gp120 gp41	Surface envelope binds to receptor molecule during viral entry Transmembrane envelope fuses with cell membrane during viral entry
tat (transactivator protein)	p14	Stimulates viral transcription; binds to TAR[a] to facilitate initiation and elongation of viral transcription
rev (differential regulator of expression of virus protein)	p13/19	Binds to RRE[b] to facilitate nuclear export of unspliced or singly spliced RNA, increases production of structural proteins
vif (virus infectivity factor)	p23	Increases virus infectivity; affects virion assembly and/or viral DNA synthesis; suppresses inhibition of viral replication by human antiretroviral factor APOBEC3G.
nef (negative regulatory factor)	p27	Pleiotropic, can increase or decrease virus replication; enhances virion infectivity, affects T-cell activation
vpr (virus protein R)	p15	Facilitates nuclear entry of preintegration complex
vpu (virus protein U)	p16	Affects virus release; disrupts Env-CD4 complexes; CD4 degradation

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

[b] RRE or Rev response element is a highly structured 351-nucleotide coding region within the *env* region. It is present in RNAs that are dependent on Rev protein for their cytoplasmic expression
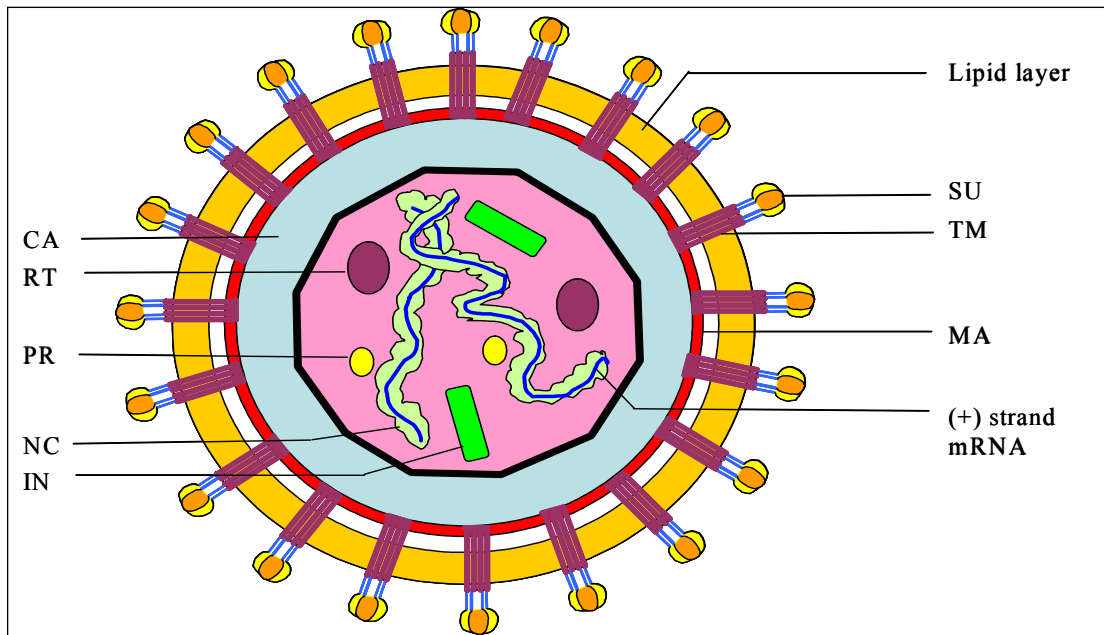
Figure 1-1: Human immunodeficiency virus

Shown is a schematic of HIV. It is an enveloped retrovirus. The envelope contains the glycoproteins gp120 and gp41 or surface unit (SU) and transmembrane (TM) protein, respectively. Within the envelope are the matrix protein (MA) and the capsid (CA) protein. The capsid encloses two copies of single strand plus sense RNA genome coated with nucleocapsid protein (NC). The capsid also contains several copies of reverse trancriptase (RT) polymerase, integrase (IN), and protease (PR) enzymes. Figure adapted from Principles of Virology, Molecular Biology, Pathogenesis and Control.

1.7 HIV genome replication

The process of reverse transcription was extensively studied by identifying replication intermediates made in endogenous or reconstituted reactions and actively infected cells [17-20]. The process takes place in a ribonucleoprotein complex that contains the diploid genome, RT, NC and some other viral proteins. RT is a multifunctional enzyme that has RNA/DNA-dependent DNA polymerase activity and RNase H activity. During replication, the polymerase activity facilitates the synthesis of DNA from an RNA or DNA template and the RNase H activity helps to cleave the RNA portion of RNA-DNA hybrid regions. NC is also a viral protein that has numerous functions during reverse transcription. It can be considered as the retroviral counterpart of prokaryotic single-stranded binding protein. It is a nucleic acid chaperone protein that aids in annealing and unwinding of nucleic acids [21, 22]. Though the two positive sense genomic RNA strands contain immediately translatable information, the virus first goes through an intermediary double stranded DNA stage *via* reverse transcription [23, 24] as shown in Figure 1-3. The two strands are joined noncovalently at multiple points; the strongest point at the 5' ends is *via* a dimer linkage structure [21, 25, 26]. The viral RNA has the structural features of a cellular messenger RNA; it includes the following from the 5' to 3': 7-me-G cap, the coding regions for *gag*, *pol* and *env* and the 3' poly (A) sequence (refer Figure 1-2).

Figure 1-2 shows the main components of the genome that participate in reverse transcription; R are the direct repeat sequences that lie adjacent to the 5' cap and the 3' poly (A) tail [27, 28], U5 and U3 are unique sequences at the 5' and 3'

9

ends, respectively, PBS is the Primer Binding Site and PPT is the polypurine tract. The overall arrangement of these elements on the viral genome from the 5' cap is: cap, R, U5, PBS, (coding regions), PPT, U3, R, poly (A). The *gag*, *pol* and *env* genes are also represented in this figure. LTRs are Long Terminal Repeats that are present only in the double stranded DNA product (Figure 1-3); they appear as a result of the duplication of the U5 and U3 sequences. The generation of the two LTRs is crucial to establish the proper DNA ends for integration. They also contain regulatory sequences (promoter and polyadenylation sites) that are important for viral transcription. The DNA complement of the + strand RNA are designated as negative (-) strands.
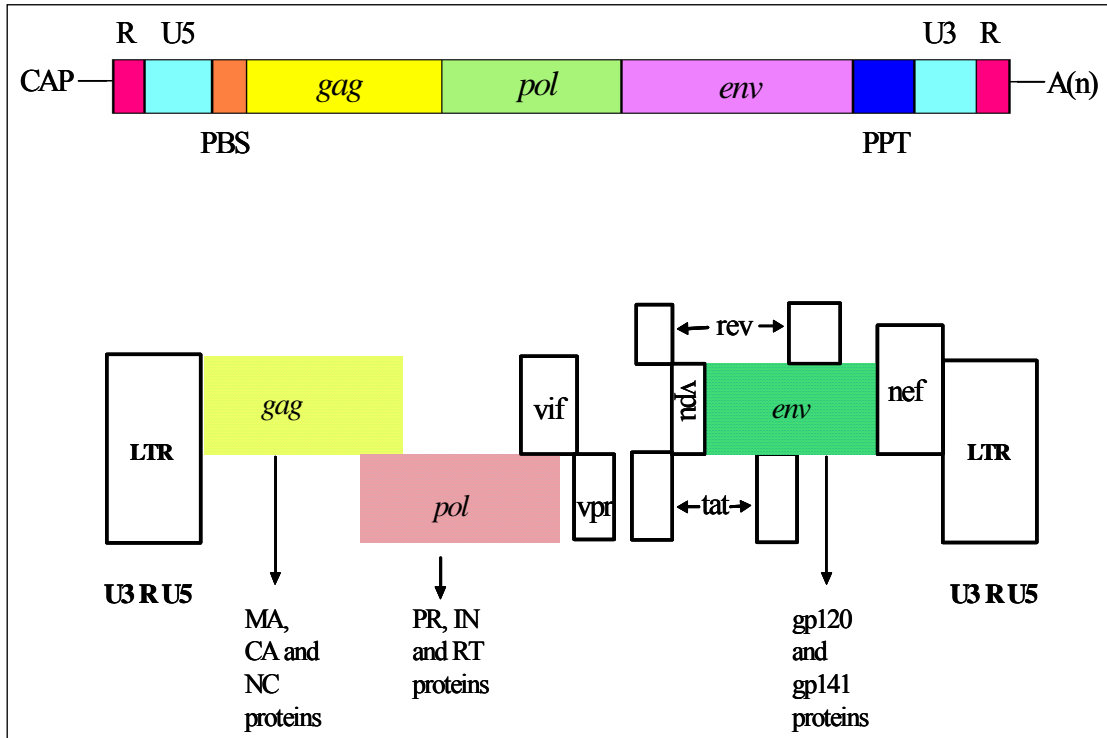
Figure 1-2: The retroviral genome and the proviral DNA

Shown are the retroviral genome and the proviral DNA. The RNA genome contains a cap at the 5' end and a polyA tail at the 3' end. PBS is the primer binding site where a tRNA primer binds to begin minus strand DNA synthesis. PPT is the polypurine tract which is the primer for the plus strand DNA synthesis. U5 and U3 are the unique sequences at the 5' and 3' ends respectively. R represents the repeat sequences at both ends. Pictured below is the proviral DNA synthesized from the RNA genome. The nine genes, their relative positions, and the positions of the LTRs or long terminal repeats are shown. The products resulting from *gag*, *pol* and *env* are described. *Tat* and *rev* are both shown upstream and downstream of *env* since splicing within *env* gene is required to complete the coding of the *tat* and *rev* proteins.

Following, is a list of the main events of reverse transcription:

a. The process of reverse transcription is triggered by a tRNA^{Lys} primer [29] that is present in the viral capsid. This primer binds to the PBS or primer-binding site near the 5$^{'}$ end of one of the RNA genomes. The RNA-dependent DNA polymerase activity of the HIV Reverse transcriptase enzyme extends this tRNA primer to produce the minus-strand DNA. When the synthesis reaches the 5' end of the genome the resultant nascent DNA is referred to as the minus-strand strong-stop DNA or -sssDNA.

b. At this stage in replication, continuation of minus-strand DNA synthesis requires the first strand transfer (also referred to as template jumping or template switching) that allows the 3' end of the genomic RNA to serve as the template [20, 30]. The nascent DNA anneals to a complementary R or repeat region at the 3' end of either the same RNA genome (intramolecular) or to the second genome (intermolecular) [31]. The missing segments of R and U5 shown on the RNA genome in Figure 1-3 indicate the RNase H activity of the RT enzyme that cleaves the RNA in RNA-DNA heteroduplex regions.

c. The extension of the minus-strand DNA by RT continues until the PBS on the RNA template and is accompanied by degradation of the template. These RNA cleavage products include a polypurine tract or PPT that serves as the primer for the synthesis of the plus-strand DNA [32]. The PPT is a highly conserved 15-base oligoribonucleotide that is resistant to RNase H degradation. The DNA dependent DNA polymerase activity of the RT enzyme synthesizes plus-strand DNA from the 3' end of PPT, until it reaches a modified base on the initial tRNA

12

primer that prevents subsequent extension. The resultant DNA product is referred to as plus-strand strong-stop DNA or +sssDNA.

d. The removal of the tRNA by RNase H activity of RT clears the way for the second strand transfer involving the +sssDNA strand. At this point in replication, the plus strand of the PBS sequences has been copied from the tRNA primer and the minus strand of the PBS has been copied from the viral genome. In the second jump, the PBS at the 3' end of the +sssDNA anneals to the PBS sequence at the 3' end of the minus-strand DNA. This transfer is wholly intramolecular and results in a circular DNA molecule with overlapping 5' ends. Now, RT finishes the synthesis of both the nascent DNA strands (both minus and plus-strand DNA) to produce a linear double stranded viral DNA molecule with long terminal repeats (LTRs). This viral DNA then randomly integrates within the host DNA where it is referred to as the provirus [10, 33, 34].

The double stranded DNA molecule then integrates into the host chromosome in a reaction catalyzed by the integrase protein. The virus now uses the host machinery for the production of viral components. HIV packages two RNA genomes into each virion. The new viral particles are released from the cell by budding.
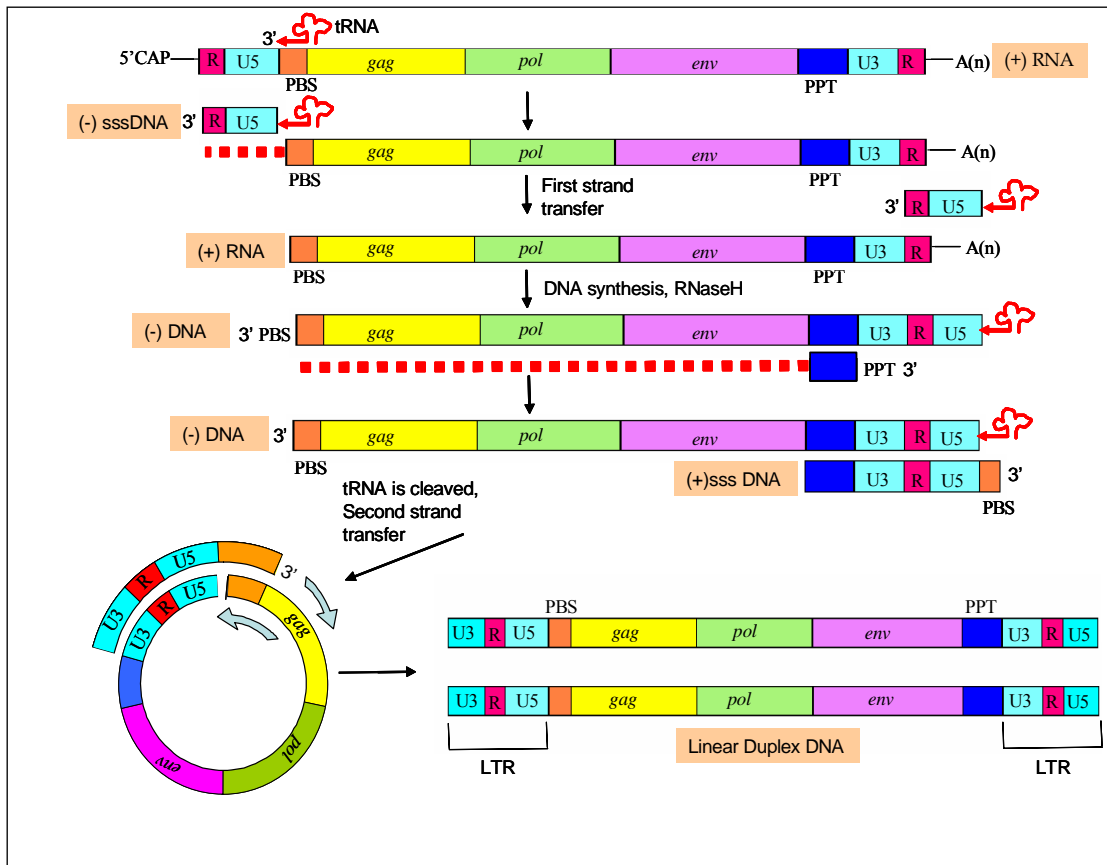
Figure 1-3: Model for retroviral replication

Shown here is the model for retroviral replication. The synthesis of proviral DNA begins from a tRNA primer that binds to the primer binding site (PBS). Synthesis of this DNA extends to the 5' end of the viral genome. This segment of DNA is called "minus strand strong stop DNA" (-sssDNA). It undergoes transfer from the 5' end to the 3' end of the same or the other RNA genome. After transfer, synthesis continues to the 5' end of the RNA. The dotted red line represents the RNA cleavage products of RNase H activity of RT enzyme. PPT or polypurine tract is the RNase H resistant segment of the genome that primes the plus strand DNA synthesis. This phase of synthesis produces a segment called "plus strand stop DNA" (+sssDNA). The second transfer involves binding of the complementary PBS regions of minus and plus DNA strands. Complete synthesis yields a double stranded DNA with long terminal repeats (LTRs).

14

1.8   Reverse Transcriptase

The hallmark of retroviral infection is the conversion of the plus strand RNA genome into a double stranded DNA, which is then integrated in the host chromosome. HIV-RT is the multifunctional enzyme primarily responsible for that conversion (Figure 1-4). It possesses an RNA-dependent DNA polymerase, a DNA-dependent DNA polymerase and also an RNase H activity that acts to degrade the genomic RNA during DNA synthesis. It exists *in vivo* as a heterodimer with two subunits of approximately 66 and 51 kDa (p66 and p51 respectively).   The p51 subunit is generated by proteolytic cleavage of the carboxyl terminus of the p66 subunit. The p66 subunit contains both the polymerase and RNase H domains, while p51 contains only the polymerase domain [35]. The role of p51 in the function of the molecule is still unclear. Some evidence suggests that the p51 subunit is not directly involved in the catalytic function of the polymerase and RNase H activities [36]. All known activities of the protein are executed by the p66 subunit of the heterodimer, as determined by studies in which the p51 subunit of the heterodimer was inactivated by mutations [35].

Both the polymerase and RNase H active site of HIV-RT consist of a catalytic core of negatively charged amino acids that serve to coordinate an essential metal cation (Mg^{2+}). Mutagenesis studies indicate that the polymerase active site includes three aspartic acid residues: Asp-185, Asp-186 and Asp-110 [37]. The RNase H active site has four conserved amino acids, Asp-443, Glu-478, Asp-498 and Asp-549 [38, 39]. HIV-RT is structurally similar to many eukaryotic and bacterial DNA polymerases, with an opene right hand conformation containing three distinct

domains (fingers, palm, and thumb) (Figure 1-4). The substrate is bound such that the primer terminus is located in the region between the thumb and fingers while the double-stranded region is located in the palm. HIV-RT displays low processivity *in vitro*, in other words, it does not remain attached to the primer-template complex for a large number of successive nucleotide additions. RT has also been found to possess very low replication fidelity with an error rate of HIV-1 RT lies between 10^{-4} and 10^{-5} per base per cycle [40, 41].
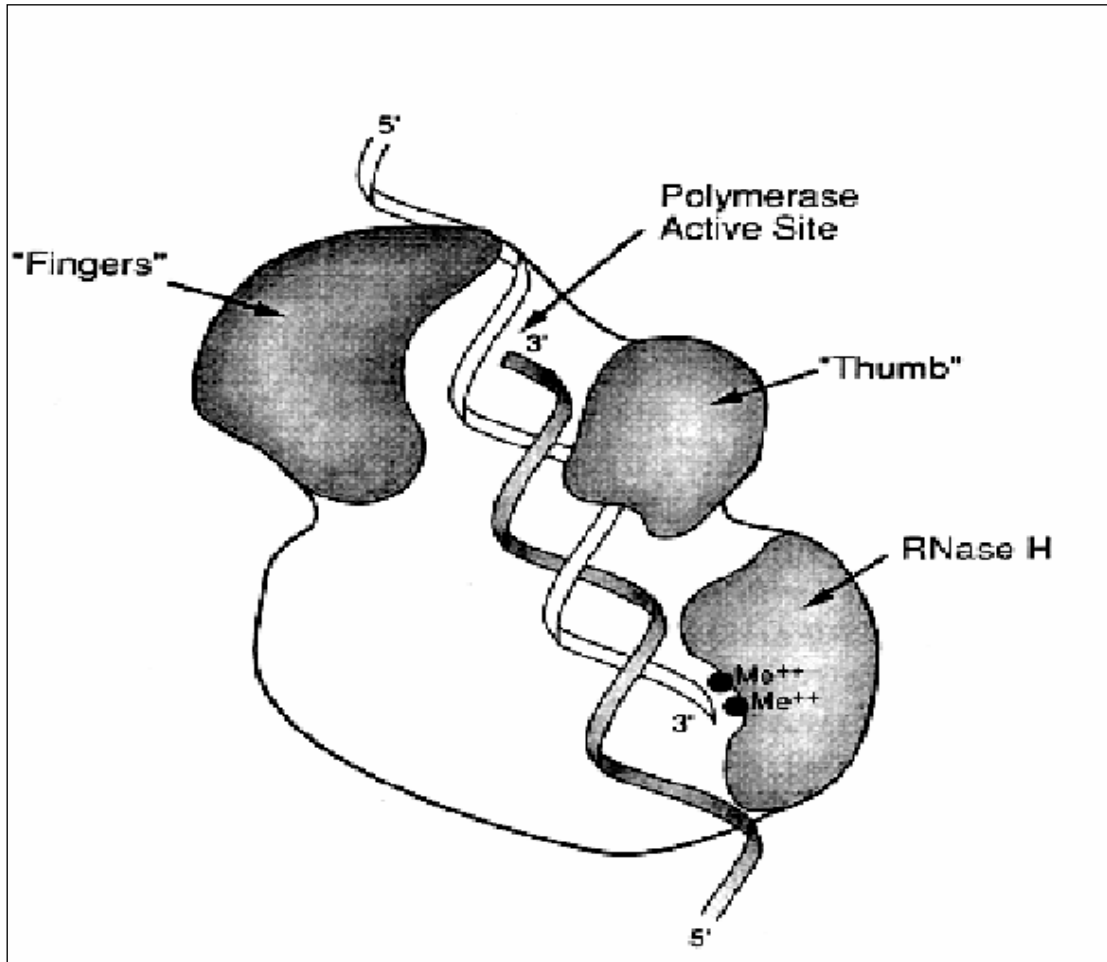
Figure 1-4: HIV reverse trancriptase

Schematic drawing of RT binding to an RNA-DNA hybrid. The polymerase active site contacts the 3' end of the DNA and RNase H active site contacts the RNA about 18 bases behind the 3' DNA end. The two Me^{++} indicate the requirement of two metal ions for the RNase H activity of the enzyme. Two metal ions are also proposed to bind at the polymerase active site. The polymerase domain is likened to a right hand leading to the naming of the polymerase subdomains as fingers and thumb. Taken from Kohlstaedt, L. A. *et al.* 1992.

1.9   Mechanisms of retroviral recombination

Genetic variation in HIV-1 is well substantiated by the large number of different HIV-1 strains isolated around the world, which have been divided into three groups; M, O and N. The major or M group has been further divided into 10 nucleotide sequence-defined subtypes [42, 43]. Genetic diversity has caused HIV-1 *in vivo* to be defined as a quasispecies, that is, a population of highly related yet genetically distinct viruses within the same individual [43].

One of the means the virus employs to generate genetic changes is recombination. Retroviruses (including HIV) copackage two RNA genomes. These genomes can be identical or genetically distinct depending on several factors. A major factor is whether the virus originated from a cell with more than one provirus in which case the RNA genomes could be generated from different sources. Second, since the genomes are ultimately produced by host RNA polymerase II the fidelity of this enzyme, which is largely unknown, would also affect the genetic make-up of each genome. HIV undergoes genetic recombinations that differ from those occurring in bacterial or mammalian systems. Retroviral recombination is *via* a process called strand transfer (also referred to as strand jumping or template switching). Strand transfer involves the switching of DNA being synthesized on one template (referred to as "donor") to homologous regions on the same or on a second template (referred to as "acceptor") where the synthesis continues. When DNA strands transfer to a different template, the resultant proviral DNA is capable of encoding a genomic RNA that is a chimera of the original parent templates. The first and second strand transfers (strong stop DNA transfers) that occur during reverse

transcription are obligatory transfers without which the viral replication cannot proceed to completion. Also, these two transfers occur at the termini of templates hence they are called terminal transfer events. In addition to these vital events, the virus can also undergo internal strand transfers which can potentially occur at any position along the genome [31, 44, 45]. These are important steps that aid in generating genetic diversity in the viral population [10, 46-48] and allow some viruses to evade host immune response and drug therapy. They also may increase the probability of successful DNA synthesis by providing a salvage pathway for broken or damaged genomes [49]. Internal transfers can occur during plus strand DNA synthesis, albeit to a lesser extent than during minus strand DNA synthesis [31]. Most transfers involve homologous regions of two genomic RNAs or the DNAs synthesized from those RNAs. The rate of nonhomologous recombination is only about 1/100 to 1/1000th of the rate of homologous recombination [50]. Jetzt *et al.* [44] have shown that on average, HIV-1 recombines approximately two to three times in every cycle of replication. More recently, Levy *et al.* [51] showed that the virus is more recombinogenic than previously known; a single round of replication in T lymphocytes in culture generated an average of nine recombination events per virus, and infection of macrophages showed about 30 crossover events per replication cycle.

There are two models that have been proposed to explain retroviral recombination; copy choice model and strand displacement-assimilation model. Experimental evidence supports both models; hence they are not mutually exclusive. The copy choice model (Figure 1-5) explains the DNA-RNA transfers that occur during the minus-strand DNA synthesis; it postulates that nascent DNA switches

from one RNA template to another during minus strand synthesis. This model is a modified version of the original 'forced copy choice' model which proposed that strand transfers have an absolute requirement for breaks on the template RNA; when RT encounters these breaks during DNA synthesis, the nascent DNA is forced to switch to homologous regions of another intact RNA template [49]. But this model was insufficient to explain all types of transfers during minus strand synthesis. For example, Hu and Temin [52] used gamma radiation to introduce breaks into RNA templates but could not demonstrate a significant increase in recombination. Efficient template switching was shown to occur from regions of unbroken RNA [53]. In undamaged RNA, secondary structures and/or sequences were shown to cause stalling or pausing of RT polymerase; which in turn induced strand transfer. Thus, the slightly modified version called copy choice model was introduced to account for all types of transfers during minus strand synthesis. The transfer of nascent DNA can occur *via* two proposed methods [54]; DNA dissociation or acceptor RNA invasion. In the first method, the nascent DNA dissociates from the original template (called donor) prior to hybridizing with a second template (called acceptor). This dissociation happens independently of the acceptor template. In the second method, the acceptor template actively displaces the DNA from the donor by invading the donor-DNA hybrid. During this invasion, there is a transient trimeric structure with the DNA bound to both donor and acceptor templates. *In vitro* [55] and *ex vivo* [56] studies have shown evidence for the existence of a trimeric structure during strand transfer. There is evidence for the occurrence of the first method also; hence, it is not

20

yet known if minus strand DNA transfers occurs preferentially through one of the two above mentioned methods.

The strand-displacement assimilation model (Figure 1-6) explains the DNA-DNA transfers during plus strand DNA synthesis [57]. Plus strand DNA synthesis is primarily initiated from the polypurine tract. But it has also been shown to be initiated at multiple alternative points [32]. During such discontinuous synthesis, the 5' ends of nascent plus strands are displaced by the 3' ends of adjacent plus strands. These strands are now free to base pair with the other minus strand DNA or incorporate with preexisting plus strand DNA fragments bound to that minus strand DNA. This model requires the presence of two minus strand DNA copies of the region where recombination will occur.
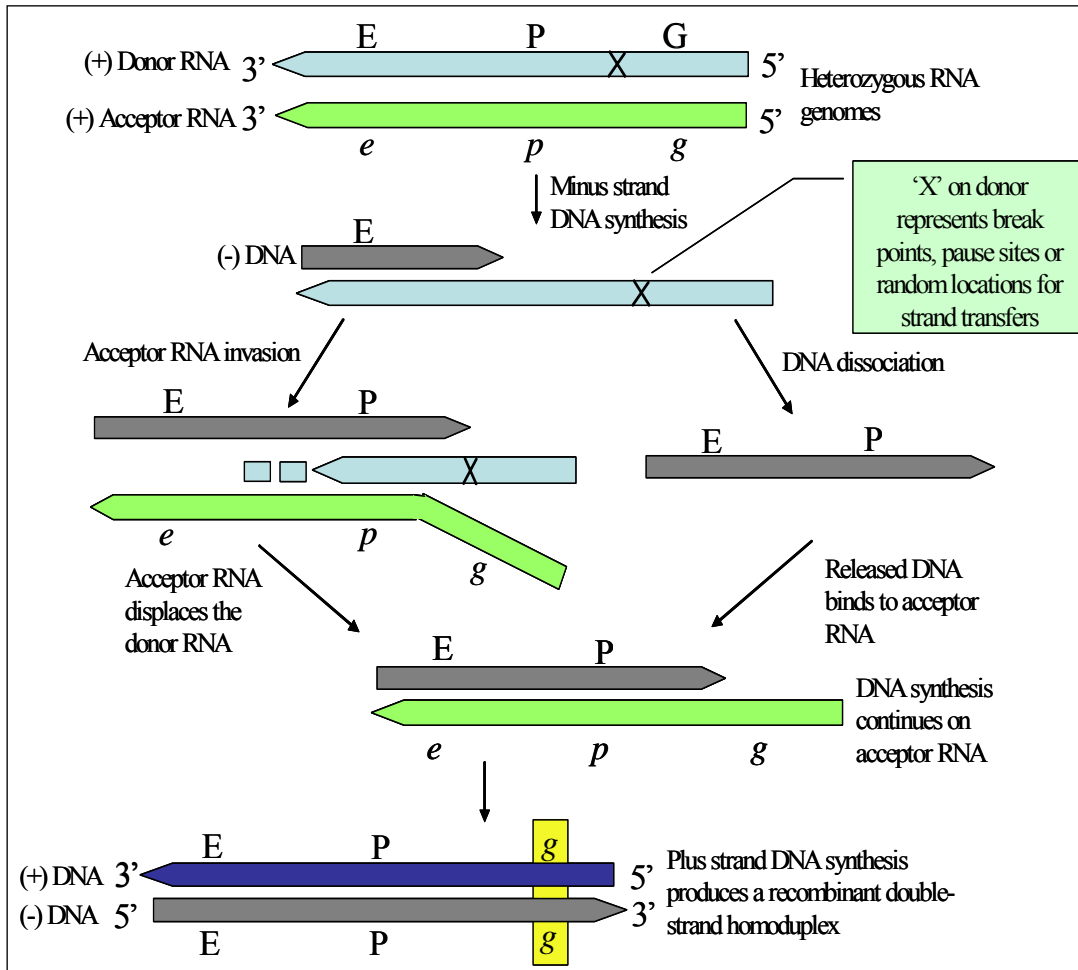
Figure 1-5: Copy choice model

Shown here is the copy choice model. This model proposes a mechanism for transfer during minus strand DNA synthesis. Viral genes *gag, pol* and *env* and their alleles are represented as G/g, P/p, and E/e respectively. Single recombination using G/g is shown for simplicity. Model includes the acceptor RNA invasion and the DNA dissociation methods of transfer proposed by several investigators. The broken blue lines represent RNA cleavage products of RNase H digestion by RT enzyme. Yellow boxes highlight the regions of recombination. Figure adapted from Principles of Virology, Molecular Biology, Pathogenesis and Control.

G   P   E

(+) Donor RNA 5'   3'   Heterozygous RNA genomes

(+) Acceptor RNA 5'   3'

g   p   e

Minus strand DNAs synthesized from both RNA templates

(-) DNA from donor RNA   G   P   E   Plus strand DNA synthesis; strand displacement of single-strand DNA tails

(+) DNA

G

(+) DNA

(-) DNA from acceptor RNA

g   p   e

Repair/Assimilation

G   p   e   Assimilation of displaced DNA produces a double-strand heteroduplex

(+) DNA 5'   3'

(-) DNA 3'   5'

g   p   e
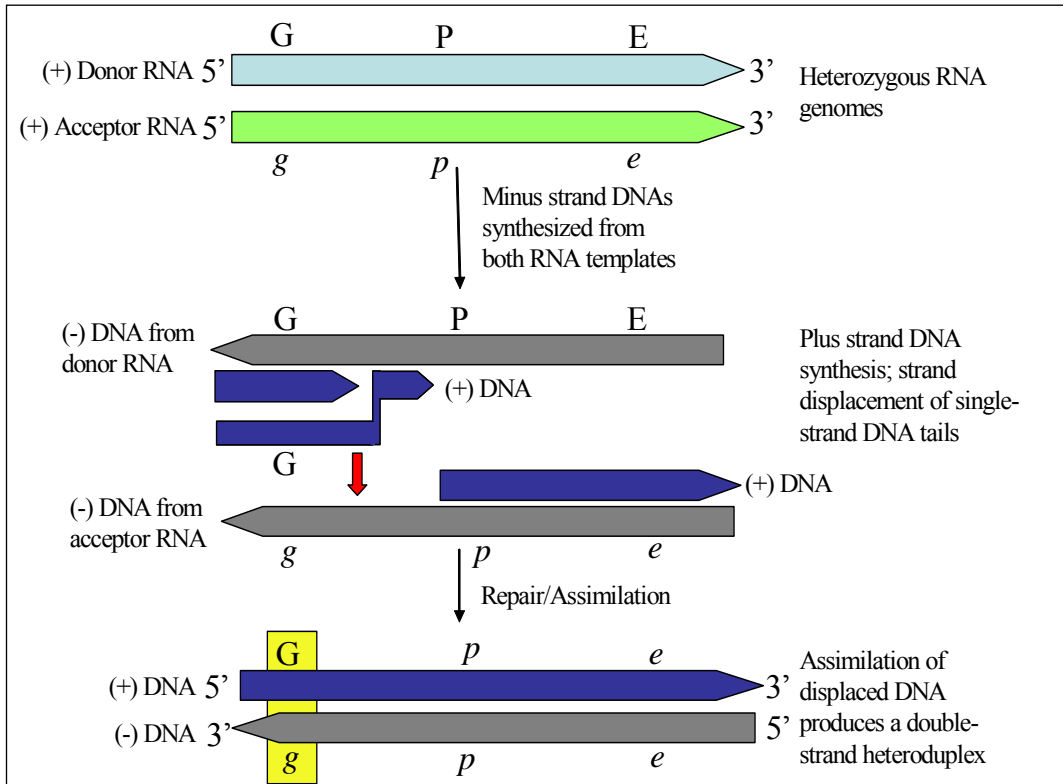
Figure 1-6: Strand displacement-assimilation model

Shown here is the strand displacement-assimilation model. This model proposes a mechanism for transfer during plus strand DNA synthesis. Viral genes *gag, pol* and *env* are shown as G/g, P/p, and E/e respectively. Single recombination using G/g is shown for simplicity. This model requires the synthesis of two minus strand DNAs. Plus strand DNA is initiated by RNA primers generated by partial RNase H digestion. The red arrow marks the point of transfer of plus strand DNA from one minus strand DNA template to the other. Yellow boxes highlight the regions of recombination. Figure adapted from Principles of Virology, Molecular Biology, Pathogenesis and Control.

1.10 Nucleocapsid protein

Nucleocapsid protein of human immunodeficiency virus type 1 is a highly basic, positively charged protein, comprised of 55 amino acid residues (Figure 1-7). All retroviral *gag* genes (including HIV) encode Gag precursor polyproteins which function in the viral assembly by recognizing and packaging two copies of unspliced viral RNA. After assembly and budding, these precursors undergo proteolysis by the viral protease to give many smaller proteins. One of the cleaved products is a small basic protein that binds single-stranded RNA and is referred to as the nucleocapsid protein [58, 59]. Initially, NC proteins were purified from virions and later on by overexpression of the protein in *Escherichia coli* strains [11]. In solution, HIV-1 NC has two rigid zinc-binding domains or zinc fingers; the two fingers are covalently linked to each other by a short flexible basic amino acid region called the linker (RAPRKKG sequence) and are flanked by flexible N- or C- terminal "tails" [60]. Each zinc fingers contains a 14-residue metal ion-binding motif, C-X$_2$-C-X$_4$-H-X$_4$-C, where X denotes variable amino acids. Deletions or alterations in the invariant CCHC motifs interfere with the specificity of viral RNA packaging and result in the absence of viral RNA in the virions [61]. The zinc fingers of NC have been shown to co-ordinate one zinc ion each, both *in vitro* and in virions [62-66]. The binding of zinc ions enables the fingers to be stable and spatially close together [67-69], however they are also susceptible to anti-viral agents that have the potential to eject zinc [70-72]. The two fingers have similar structures [73] but are not functionally equivalent [74, 75]. The NC proteins from HIV strains vary somewhat but have highly conserved hydrophobic residues at positions 13, 16, 24 and 25; a basic residue at

position 26 and a high frequency of glycine residues at positions 19 and 22 [76-78].

NMR methods have been used to determine the three-dimensional structure of NC to a high atomic resolution [76]. NC binds to various nucleic acid targets during the viral life cycle; consequently, there are no established functional assays to determine the sequence or tertiary structure of preferred targets. Hence, most studies on NC's interaction with nucleic acids have utilized fluorescent nucleic acids, RNA fragments as probes etc. De Guzman *et al.* [79] have conducted in depth analyses of NC interaction with the stem-loops (SL1-SL4) that form the 120-nucleotide Ψ-packaging signal at the 5' end of unspliced viral RNAs. This site is recognized by the NC part of the Gag polyprotein during viral assembly. In the NC-SL3 complex, the amino acid residues Lys3-Arg10 of the N-terminal tail (Figure 1-7) form a 3_{10} helix that binds within the major groove of SL3, while the zinc knuckles interact with the exposed bases, specifically the guanosines of the RNA stem. Side chains of Val13, Phe16, Ile24 and Ala25 residues form a hydrophobic cleft where hydrogen bonds between backbone amide atoms of Phe16 and Ala25 and guanosine residues occur. Arg32 forms electrostatic interactions with the phosphodiester backbone of the nucleic acid. Similarly, a hydrophobic cleft is formed by the side chains of residues Trp37, Gln45, and Met46. NC protein and SL2 of the Ψ-site exhibit similar interactions as the SL3 with a few exceptions; the helix binds within the minor groove of SL2 instead of the major groove and the hydrogen bonds are substituted with electrostatic interactions.
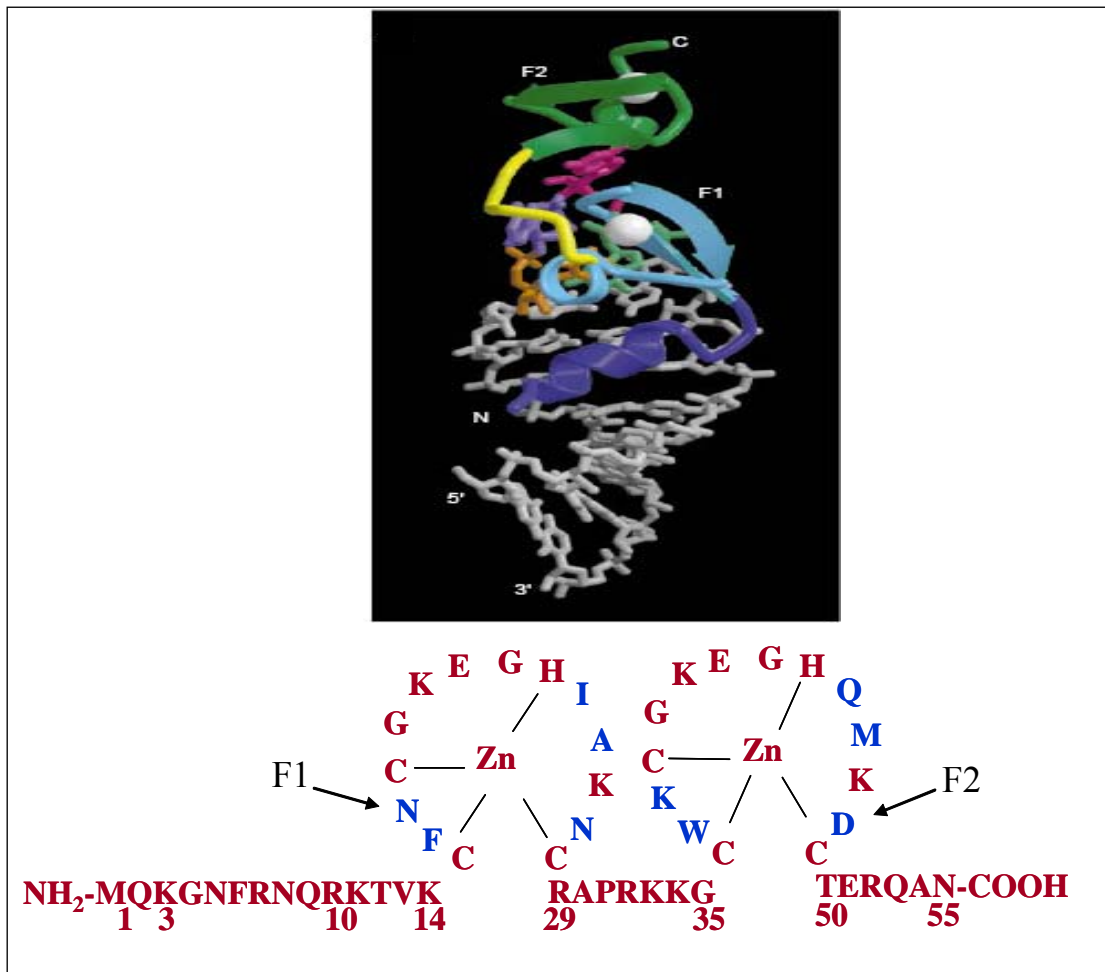
Figure 1-7: HIV nucleocapsid protein

Shown above is the ribbon diagram of the HIV-1 NC-SL3 Ψ-RNA complex. Color code: 3_{10} helix, purple; first zinc finger, blue; linker segment, yellow; second zinc finger, green; zinc atoms, white spheres; RNA, gray, except for the guanosine residues which are colored (figure taken from De Guzman *et al* [79]). Shown below is the 55 amino acid sequence (numbered 1-55) of HIV NC protein, with amino and carboxyl terminals (figure adapted from Rein *et al* [60]). There are two zinc fingers F1 and F2, each of which has the C-X_2-C-X_4-H-X_4-C invariant motif, where X denotes the variable amino acids. The two fingers differ from each other by five amino residues (highlighted in blue).

NC exhibits a multitude of essential functions in the life cycle of HIV [65, 80-83] that make it an ideal target for drug therapy [84-86] and vaccine development [87]. It binds to both RNA and DNA with a preference for single-stranded molecules; it binds to RNA in the following order of affinity: retroviral RNA > mRNA > rRNA > poly(rA) [65, 88]. At saturating levels, NC protein covers nucleic acids and offers an incomplete protection from nuclease attack [89]. All the functions of NC are credited to a remarkable biochemical activity; 'nucleic acid chaperone' activity, which is the ability of NC protein to catalyze the rearrangement of a nucleic acid molecule into an optimum conformation that has the maximal number of base pairs [60]. The exact mechanism that triggers this activity is not completely understood. Upon interacting with a nucleic acid molecule, NC protein causes transient unpairing of bases. These free bases are now available for re-pairing in optimal conformations. This phenomenon allows nucleic acid molecules to escape from suboptimal conformations. The chaperone activity is manifest in a wide variety of functions of NC protein, for example, strand exchange, annealing and unwinding functions. NC protein can facilitate transfer of a nucleic acid strand from a less stable hybrid to a more stable hybrid [90], accelerate annealing of complementary strands and cause unwinding of structured RNAs like tRNA and the TAR stem loop on the viral genome [65, 91]. A number of other nucleic acid binding proteins have been shown to facilitate melting and annealing reactions in a similar fashion. *E. coli* single-stranded DNA-binding (SSB) causes a reduction in the electrostatic repulsion of negatively charged DNA, melts out intramolecular structures and hence enhances

renaturation [92]. RNA binding proteins like heterogeneous nuclear ribonucleoprotein A1 also accelerate annealing and strand exchange [93].

A preponderance of reports has detailed the involvement of NC protein in almost every step of the HIV-1 replication cycle. In the initial stages of the reverse transcription the 3' 18 nucleotides of human tRNA^{Lys,3} are annealed to a complementary sequence on the viral RNA genome called PBS or primer binding site; NC protein facilitates the unwinding [82, 94] and annealing of the tRNA^{Lys,3} primer [95-97]. During reverse transcription, NC protein enhances strand transfer (a process that leads to recombination) [98-100]. There are two obligatory strand transfers that occur at the termini of the viral RNA [46, 48, 49]. NC has been shown to enhance these two obligatory transfers [98, 100-102]. The first strand transfer involves the minus strand strong stop DNA (-sssDNA) that is synthesized from the tRNA primer annealed to the 5' end of the genome. The -sssDNA must be translocated to the R region at the 3' end of the genome for continuation of synthesis. In *in vitro* assays, the level of this first transfer event rises from 3% in the absence of NC to as much as 65% in the presence of high levels of the protein [60]. The second strand transfer involves the plus strand strong stop DNA (+sssDNA) that is synthesized from a polypurine tract primer. It is made by copying the minus strand DNA sequences including the 18 nucleotides at the 3' end of the tRNA primer that annealed to the PBS on the viral genome. For continued synthesis, the tRNA primer must be removed so that PBS sequences at the 3' ends of the plus and minus strand DNA can anneal. NC protein has been shown to catalyze tRNA removal and annealing of the PBS sequences [103].

In addition to the above mentioned terminal transfer events the nascent DNA can switch templates at any internal position along the genome; this is referred to as internal strand transfers and is the mechanism by which copy-choice type recombinations occur. It has been shown that some secondary structures and/or sequences on the viral genome pose as barriers in the path of RT enzyme; they are called 'pause sites' as they can potentially cause pausing of RT during DNA synthesis. At these sites synthesis can be rescued if the stalled DNA transfers to the other viral genome present in the virion. NC has been shown to reduce pausing and promote strand transfer from pause sites [80, 99, 104, 105]. The mechanism by which NC protein promotes these processes is not yet completely understood but presumably stems from a combination of the nucleic acid annealing and helix destabilization/unwinding activities of NC [90, 99, 100, 106-108]. Work done in this thesis sheds some light on the effects of NC protein on internal strand transfers from various regions along the viral genome.

Upon completion of reverse transcription, NC is possibly involved in integration of the double stranded DNA into the host genome [89]. During viral assembly, NC and all the other Gag and Pol proteins exist as domains of a large polyprotein. NC in the polyprotein form is capable of a highly specific interaction with an intact packaging signal on the viral RNA [109]. Hence, NC also participates in recognition and packaging of the viral genome [75, 110-112]. In a budding virion, the genomic RNA is actually present as a dimer of two nearly identical, positive-strand RNAs that are held together *via* a limited number of base pairs. In a mature virus, the dimer undergoes a conformational change called maturation whereby it is

converted into a more compact and more thermo-stable dimeric state. NC protein has been implicated in the maturation of genomic RNA dimer [113-115]

The crucial functions carried out by NC protein during replication and infection has placed it in the same league as the other three classical retroviral proteins – RT, protease and integrase – as a potential target for drug therapy and vaccine development. The high frequency of recombination in HIV-1 has accelerated the generation of multidrug resistance virions. The recombination rates of HIV are higher than all other retroviruses thus far tested; for example, Rhodes *et al.* showed that HIV-1 recombines 10-fold higher than spleen necrosis virus and murine leukemia virus [116]. A thorough examination of factors that influence recombination is important to identify new targets for drug therapy and vaccines. Work done in this thesis examines HIV-1 recombination in the context of internal strand transfer events. It also examines the effects of various factors like NC protein, 'pausing' by RT enzyme and RNA structural intricacies on recombination**.**

Chapter 2   Differential effects of NC protein on strand transfer in various regions of the HIV genome.

2.1 Introduction

Internal strand transfers within the viral genome have been shown to occur at a high frequency, especially during minus strand DNA synthesis [46, 117, 118]. The role of pausing by RT enzyme during these events has been studied extensively by several investigators [54, 119, 120]. The progress of RT along the genome can be hampered by many sequences and structures that subsequently cause a break in continued synthesis. Such positions are called pause sites. They can be viewed as serving the role of break points on undamaged RNA. The pause-driven transfer mechanism proposes that the paused DNA products are the substrates that ultimately are used to produce transfer products. This mechanism involves a complex interplay of nucleic acids, nucleocapsid protein and RT enzyme. The main functions of HIV RT are; polymerase (can synthesize on RNA and DNA templates) and RNase H activities. During synthesis RNA-DNA hybrids are formed and the RNase H activity of RT specifically cleaves the RNA portion of these hybrids. Previous reports have shown that the polymerase and the RNase H activities of RT are not strictly coupled [121] and the rate of polymerization is generally much faster than RNase H [122]. This means that during uninterrupted polymerase activity, the RNA template is not extensively degraded. This situation changes at a pause site; here the RNase H activity takes precedence over synthesis [123]. Degradation of the RNA template destabilizes the donor RNA template-nascent DNA and also clears regions on the DNA that are free to bind to the acceptor template. The nascent DNA may be

released from the donor, or the acceptor RNA may invade and displace the donor RNA (Figure 1-5). Eventually, the nascent DNA associates with the acceptor RNA, where synthesis continues. The manifold effect of NC protein on pause mediated transfers has been well studied [80, 102, 104]. The main components of NC's nucleic acid chaperone activity are annealing and melting of nucleic acids. At pause sites, NC would be expected to melt out the obstructing secondary structures to facilitate continued synthesis on the donor template. Strand transfer experiments on a secondary structure from the polypurine tract of murine leukemia virus showed that NC did indeed reduce pausing and stimulate continued synthesis of nascent DNA on donor template [107]. But in the presence of acceptor template, NC resolves pause sites with an accompanying increase in strand transfer [124]. It has been proposed that NC's effect on pause mediated transfer is *via* the following steps; (i) it causes enhancement of RNase H activity of RT and promotes rebinding of templates by RT, (ii) it promotes the strand exchange and annealing of the nascent DNA from the degraded donor RNA to the intact acceptor RNA, (iii) it melts out secondary structures on various nucleic acids so they can bind together easily.

A preponderance of reports supports the above described effects of NC on both internal and terminal strand transfers [65, 101, 106]. Pause mediated transfers have been observed in a variety of templates; for example, elimination of a pause site in the HIV-1 *nef* gene with minimal sequence change resulted in a drop in strand transfer [125], pause sites were detected in the TAR hairpin structure that participates in the first strand transfer [124] etc. Yet, it is not known if pause mediated transfers are the driving force for all recombination events. Some investigators have proposed

a pause independent mechanism for strand transfer where the acceptor RNA structure plays a predominant role [126]. A number of hotspots for recombination have detected in cell culture infection assays (*ex vivo* assays) studies and it is not known if they all correlate with pause sites [31]. Also, though a majority of earlier reports on strand transfer have elaborated upon the activity of NC in highly structured RNA templates, relatively little is known about the effects on RNA templates with weak or no structure. This report has addressed some of these questions: (i) Presuming that strand transfers occur all along the genome, does NC exhibit a consistent or a differential influence on strand transfer along the length of the viral genome? For example, in regions with strong secondary structures, NC may be required to unwind these structures to facilitate association of the nascent DNA and acceptor RNA, whereas in regions relatively devoid of stable structures, NC may exert only a minimal influence on strand transfer, (ii) Do *ex vivo* hotspots for recombination enhance strand transfer *in vitro* also? Do they all contain strong pause sites that induce strand transfer? (iii) Is the extent of *in vitro* strand transfer different in highly structured *vs.* weakly structured regions of the viral genome. To address these questions, strand transfers in five different regions of the HIV genome were analyzed; three of these regions are highly structured while two have relatively weak structures. The highly structured regions consisted of the following; a region containing a portion of the U3 3'LTR, the *gag-pol* frameshift region and the Rev response element (RRE) region. The first one was detected as a recombination hotspot by Yu *et al*. [31] in cell culture assays and the last two are important HIV regulatory sites that also have a high potential for pause sites. The two weakly structured regions are from the

pol-vif and *env* regions; they have low potential for pause sites. They were chosen after scanning several regions of the genome using RNA folding programs to predict folding stability. These studies were expected to reveal the global effects of NC protein on the HIV-1 genome and the importance of RNA structural intricacies on strand transfer.

2.2 Materials

Plasmid pNL4-3 obtained from the NIH AIDS Research and Reference Reagent Program contains a complete copy of the HIV-1 provirus derived from strains NY5 and LAV [127]. PCR primers were obtained from Integrated DNA Technologies, Inc. Recombinant HIV-RT was graciously provided to us by Genetics Institute (Cambridge, MA). This enzyme has a specific activity of about 40,000 units/ml (one unit of RT is defined as the amount required to incorporate 1 nmol of dTTP into nucleic acid product in 10 minutes at 37°C using oligo(dT)-poly(rA) as primer-template). The enzyme contained very low levels of single-stranded nuclease activity, which was found to be inhibited by including 5mM AMP in the assays [54]. AMP at this concentration did not affect the polymerase and RNase H activities of RT. Aliquots of HIV-RT were stored frozen at -70°C, and a fresh aliquot was used for each experiment.

HIV-NC clone was a generous gift from Dr. Charles McHenry (University of Colorado). NC was purified according to the protocol described [11]. The purity of the protein was evaluated using coomassie blue staining of 17.5% SDS-PAGE gels [128]. Quantification was by absorbance at 280 nm using an extinction coefficient of 8350 cm-1M-1 [11]. Aliquots of NC were stored frozen at -70°C, and a fresh aliquot was used for each experiment. Taq polymerase was from Eppendorf. SP6 RNA polymerase, DNase I-RNase free, and RNase-DNase free were from Roche Diagnostics. RNase inhibitor was from Promega. T4 Polynucleotide Kinase was obtained from New England Biolabs. Proteinase K was obtained from Kodak.

Radiolabeled compounds: γ 32P ATP was obtained from Amersham. Sephadex G-25 spin columns were from Amika Corp. All other chemicals were from Sigma or Fisher Scientific.

2.3 Methods

PCR amplification of DNA substrates: Five different sets (4 primers per set) of PCR primers were specifically designed to yield donor and acceptor RNAs that shared homology over a 150 base region. The primers amplified DNA from five different areas of the pNL4-3 plasmid (Table 2-1) (Figure 2-1). An SP6 promoter sequence was included on some of the primers to allow transcription of the DNA by SP6 RNA polymerase. PCR reactions were performed according to the enzyme manufacturer's protocol using the provided buffer. One hundred pmoles of each primer was used. The following cycling parameters were used; 35 cycles of denaturation, annealing and extension at temperatures of 94°C, 50°C and 72°C, respectively, for 1 min each, followed by one cycle of extension at 72°C for 5 min. The PCR products were purified on 8% native polyacrylamide gels (29:1 acrylamide: bisacrylamide) and used to prepare RNA as described below. The DNAs were excised and eluted overnight in a TE buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)). The eluate was separated from the gel by centrifugation and subsequent filtration through a 0.45µ disposable syringe filter. The DNAs were recovered by precipitation in ethanol with 300 mM sodium acetate (CH3COO.3H2O). The amount of recovered DNAs was determined spectrophotometrically from optical density.

Preparation of RNA substrates: Run-off transcription (performed according to the enzyme manufacturer's protocol) was conducted using approximately 5 µg of the purified PCR DNAs and SP6 RNA polymerase enzyme to generate donor RNA transcripts of 175 nucleotides and acceptor RNA transcripts of 177 nucleotides. The transcription reactions were treated with 0.4-units/µl of DNase I-RNase free enzyme

for 15 min to digest away the template DNA. Then they were extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with ethanol. The RNA pellets were resuspended in 50 μl of RNase-free water, loaded onto two successive hydrated sephadex G-25 spin columns and processed according to the manufacturer's directions. The amount of recovered RNAs was determined spectrophotometrically from optical density.

RNA-DNA hybridization: DNA primers that bound specifically to the donor RNA transcripts were ^{32}P-labeled at the 5' end with T4 polynucleotide kinase according to the manufacturer's protocol. Each of the five donor RNAs was hybridized to a complementary labeled primer by mixing primer:transcript at approximately 5:1 ratio in 50 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 80 mM KCl, and 0.1 mM EDTA (pH 8.0). The mixture was heated to 65°C for 5 min and then slowly cooled to room temperature.

Time course reaction: Donor RNA-primer DNA hybrids (2 nM final concentration of RNA) were preincubated for 3 min in the presence or absence of 10 nM acceptor RNA template and NC (as indicated) in 42 μl of buffer (see below) at 37°C. One molecule of NC per two nucleotides was used in the reactions, i.e. final concentrations of 1.16 μM and 0.27 μM of NC were used in reactions with or without acceptor RNA templates, respectively. The reactions were initiated by the addition of 8 μl of HIV-RT at a final concentration of 2.5-units/μl. The following reagents at the indicated final concentrations were also included in the reaction mixtures: 50 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 80 mM KCl, 0.1 mM EDTA (pH 8.0), 6 mM

MgCl2 , 100 μM dNTPs, 5 mM AMP (pH 7.0), 25 μM ZnCl2 and 0.4-units/μl RNase inhibitor. Reactions were allowed to incubate for time points of 0, 2, 4, 8, 16, 32, and 64 min at 37°C. At these time points, a 6 μl aliquot of each reaction was terminated by mixing with 4 μl of a solution containing 25 mM EDTA (pH 8.0), 2.5 ng of RNase-DNase free enzyme and allowed to digest for 20 min at 37°C. Two μl of proteinase K at 2 mg/ml in 1.25% SDS, 15 mM EDTA (pH 8.0) and 10 mM Tris (pH 8.0) was then added to the above mixture, which was placed at 65°C for 1 hour. Finally, 12 μl of 2X formamide dye (90% formamide, 10 mM EDTA (pH 8.0), 0.1% xylene cyanol, 0.1% bromophenol blue) was added to the mixture and the samples were resolved on an 8% denaturing polyacrylamide gel containing 7 M urea. Extended DNA products were quantified by phosphorimager analysis using a GS-525 phosphorimager from Bio-Rad.

NC titration experiments: In the NC titration experiments the total reaction volume was reduced to 12.5 μl and the amount of NC in the reactions was varied as indicated in the Figure 2-12 and the reactions were allowed to proceed for 32 mins. Extended DNA products were quantified by phosphorimager analysis using a GS-525 phosphorimager from Bio-Rad.

Gel electrophoresis: Denaturing 8% polyacrylamide gels (19:1 acrylamide:bis-acrylamide), containing 7 M urea, and native 8% polyacrylamide gels (29:1 acrylamide:bis-acrylamide) were prepared and subjected to electrophoresis.
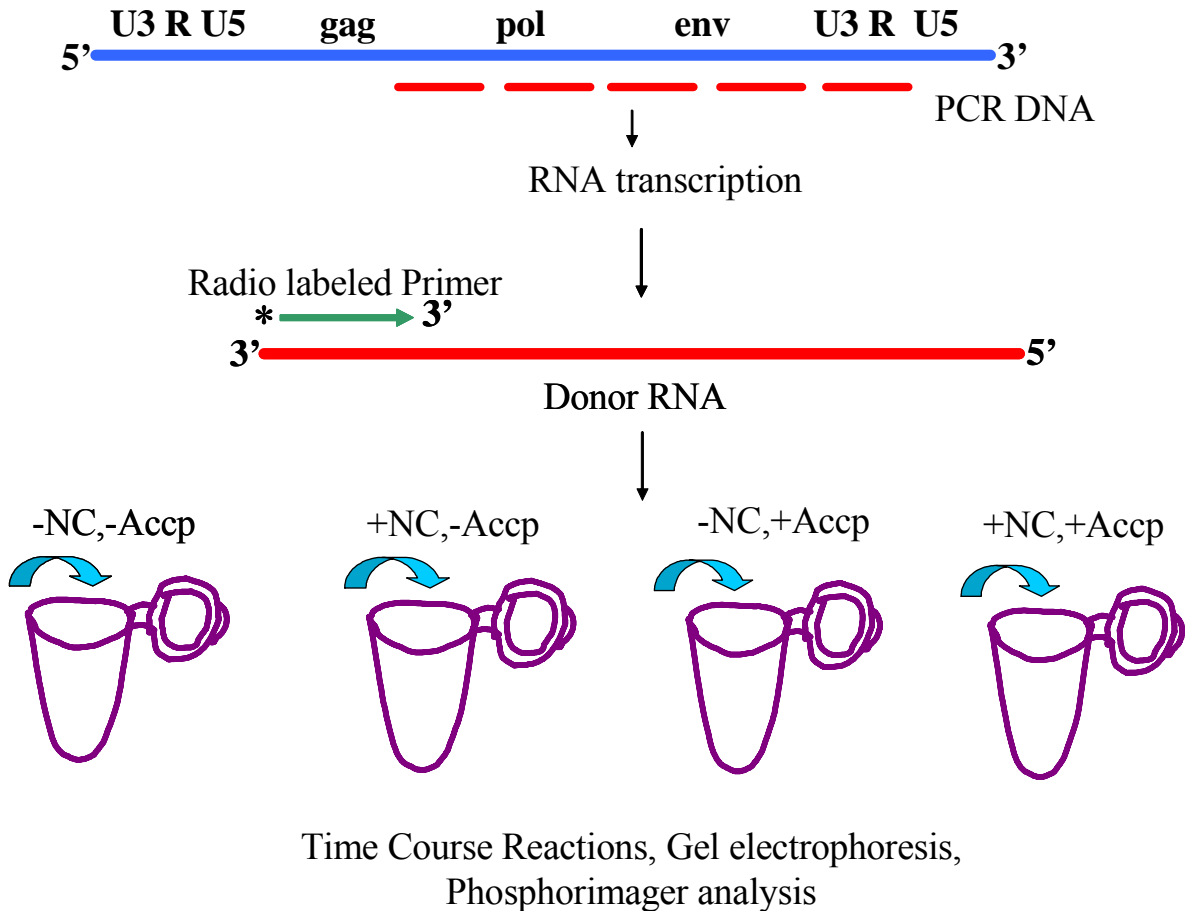
Figure 2-1: Experimental approaches

Shown above is a schematic of experimental approaches. PCR DNA (five short red lines) was amplified from different regions on the pNL4-3 plasmid (blue line). RNA transcription was conducted to get donor and acceptor RNA templates. Only the donor RNA is shown as a red line. Radio labeled primers (green line with asterisk) synthesized DNA on the donor. The primer-donor RNA hybrid was used in four different reactions (as shown). –NC and +NC represents the absence and presence of nucleocapsid protein, respectively. –Accp and +Accp represents the absence and presence of acceptor RNA, respectively.

Table 2-1: PCR primers used in the amplification of various segments of the HIV-1 genome

REGION	BASE NUMBERS[A]	PRIMER SEQUENCES[B]
U3 3'LTR donor	9096-9265	a. 5'**gatttaggtgacactatag**_atata_caaagaagacaagatatcct 3' b. 5' tctcctttattggcctcttc 3'
U3 3'LTR acceptor	9074-9245	a. 5' **gatttaggtgacactatag**_tatat_actggaagggctaattcact 3' b. 5' taccttatctggctcaactg 3'
PolVif donor	3441-3610	a. 5' **gatttaggtgacactatag**_atata_gaagcagagctagaa 3' b. 5' tattttcctgttttcagatt 3'
PolVif acceptor	3419-3590	a. 5' **gatttaggtgacactatag**_tatat_agaagtaataccactaacag 3' b. 5' tttaaatggctcttgataaa 3'
GagPol donor	2031-2200	a. 5' **gatttaggtgacactatag**_atata_gaaatgtggaaagga 3' b. 5' ttgttgtctctaccccagac 3'
GagPol acceptor	2009-2180	a. 5' **gatttaggtgacactatag**_tatat_cccctaggaaaaagggctgt 3' b. 5' ctgaagctctcttctggtgg 3'
Env donor	7101-7270	a. 5' **gatttaggtgacactatag**_atata_gtacaagacccaaca 3' b. 5' ttgttctcttaatttgctag 3'
Env acceptor	7079-7250	a. 5' **gatttaggtgacactatag**_tatat_aacacatctgtagaaattaa 3' b. 5' ctatctgttttaaagtggca 3'
RRE donor	7823-7992	a. 5' **gatttaggtgacactatag**_atata_ctgacggtacaggcc 3' b. 5' aggagctgttgatcctttag3
RRE acceptor	7801-7972	a. 5' **gatttaggtgacactatag**_tatat_tatgggctgcacgtcaatga 3' b. 5' gtatctttccacagccagga 3'

A-Refers to sequence numbering of plasmid pNL4-3 that contains an almost complete copy of the HIV-1 provirus derived from strains NY5 and LAV.

B-The primer sequences used to prime the pNL4-3 plasmid. The bases in bold face are the SP6 promoter sequences and those in italics are non-retroviral sequences that were added to prevent end transfers. All the a. primers are in 5' orientation and the b. primers are in 3' orientation.

2.4 Results

Strand transfer assay: The general approach used to test for strand transfer in the different regions of the HIV genome is depicted in Figure 2-1 and Figure 2-2 . This assay is designed to simulate internal strand transfer events occurring during minus strand DNA synthesis. The donor (template on which DNA synthesis initiates) and acceptor (template to which DNAs initiating on the donor can potentially transfer to) represent the two strands of viral genomic RNA in the virion. DNA synthesis is initiated from a 5' end-labeled DNA primer that was specifically designed to bind only to the 3' end of the donor RNA. Strand transfer can occur over the transfer zone, which is the region of homology between the donor and acceptor RNAs. Primer extension to the end of the donor produces a 175-base full-length donor-directed DNA product. Strand transfers result in 197-base transfer DNA products that include the additional non-homologous bases at the 5'end of the acceptor RNA. The 5' end of the donor outside the transfer zone is not homologous to the acceptor, which prevents DNAs extended to this region from transferring. This limits the system to analysis of internal transfer events between the donor and acceptor rather than events occurring from the template termini. The difference in the lengths of the two DNA products allowed quantification of strand transfer events in donor-acceptor pairs from five different regions of the HIV genome (refer Figure 2-3 and Figure 2-4). It is also possible for strand transfer to occur between two donor RNAs in the reaction. This type of transfer was not quantified in the experiments. However, it probably represents only a small amount of the total transfer, since reactions included

a 5-fold excess of acceptor over donor and also, many of the donors would not be available since they are used for DNA synthesis.

The percent transfer efficiency of strand transfer was defined as ((Transfer DNA products (T))/(Transfer + Full-length donor-directed products (F))), times 100 ((T/(T+F)) x 100). The number reflects the proportion of DNA primers extended to the end of the acceptor versus those extended to the end of the donor. This representation of the data, as opposed to simply determining the gross level of transfer product, expresses transfer relative to total DNA extension. Therefore, differences in the total amount of primers extended with the various substrates used are compensated for.

Homologous Transfer Zone (150 nt)

*5' —————————▶ 5' Donor (175 nt)

————————————————— (red line)

————————▶ 5' Acceptor (177 nt)

————————————————— (red line)

(175 nt)
Full- length
donor-directed DNA

*5' ————————————————▶ Transfer DNA
(197 nt)
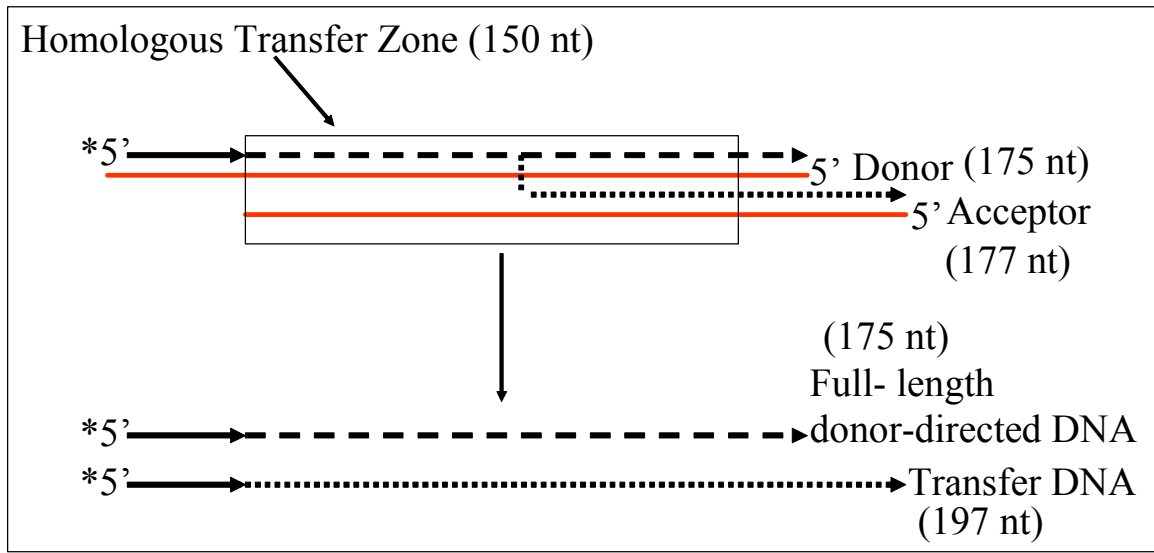
Figure 2-2: Strand transfer assay

Schematic representation of the strand transfer assay is shown here**.** Donor and acceptor RNA templates are labeled along with their corresponding total lengths. The 150 nucleotide (nt) boxed region enclosing the two templates is the region of homology or the transfer zone. The asterisk represents the 5' labeled primer that is complementary to only the donor template. The broken lines represent the full length DNA that is synthesized on the donor and the dotted lines represent the transfer DNA that has undergone a strand transfer event to the acceptor. The lengths of full-length donor-directed DNA and transfer DNA products are indicated at the bottom of the figure

Prediction of secondary structures of the RNA substrates: RNAdraw [129] (http://iubio.bio.indiana.edu/soft/molbio/ibmpc/rnadraw-readme.html) and Zuker's RNA folding programs [130-132] (http://www.bioinfo.rpi.edu/applications/mfold/) were used to predict the secondary structures for the RNA substrates (both donor and acceptor) utilized in this report. Figure 2-3 and Figure 2-4 show the structures of only the donor RNAs derived from the U3 3'LTR (2A), *gag-pol* (2B, referred to as GagPol substrate), RRE (2C), *pol-vif* (2D, referred to as PolVif substrate) and *env* (2E, referred to as Env substrate) regions. In each region, the acceptor RNAs were predicted to have structure similar to the donor (acceptor structures not shown). A highly negative ΔG value and a high base pair melting temperature indicated the presence of strong structures. For example, the donor GagPol RNA had a predicted ΔG value of -45.3 kcal/mole and the predicted stem loops persisted even at temperatures above 55^0C whereas the donor Env RNA showed a $\Delta G= -15.0$ and the structure melted out completely above 55^0C. Although the RNA fold programs may not predict the structure of RNAs with 100% reliability, a reasonable estimate of the strength and characteristics of the RNA would be expected, especially since relatively small RNAs were used. In addition, the GagPol and RRE regions used have previously been shown to possess strong stem loop structures similar to those predicted by the folding programs [133-136]. For our purposes an exact rendering of the structures is not necessary, just a general idea of their relative strengths and the presence or absence of pause site is required. The latter can be evaluated from RT primer extension reactions.
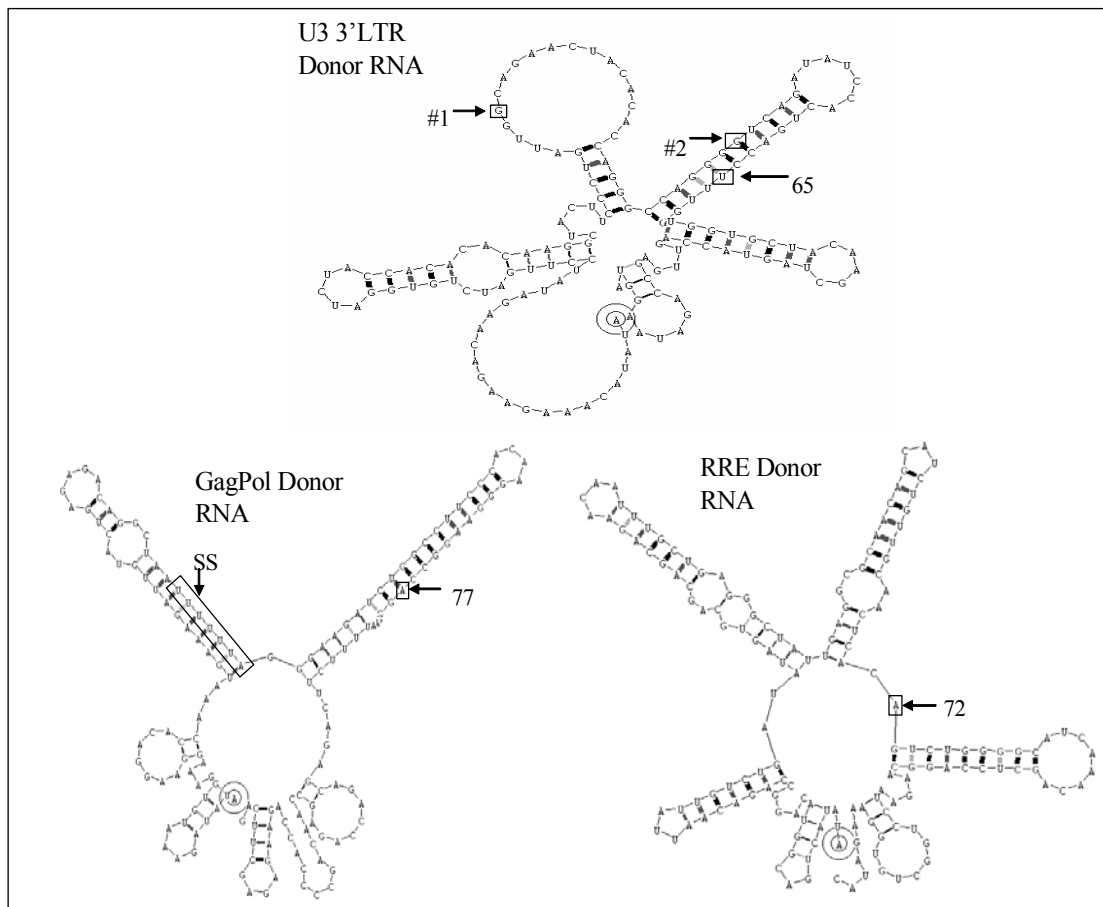
Figure 2-3: Highly structured templates

Predicted secondary structures of highly structured donor RNA templates. The base encircled in two rings is the 5' end. The probability of formation for particular base pairs is indicated by the thickness of the lines between them. Pause sites for each template are shown here (explained in Results). U3 3'LTR donor template is shown (bases 9096-9265 and ΔG = -49.30 kcal/mole). The number 65 marks a major pause site at a U residue (enclosed in the box) that is 65 nt from the 3' end of the donor template (note that the 20 nt binding to the primer are not shown). The #1 and #2 markings are the boundaries of the 24 nt potential hotspot (see Results). GagPol donor template is shown (bases 2031-2200 and ΔG = -45.30 kcal/mole). The number 77 marks a major pause site at an A residue. SS refers to the slippery site heptamer sequence. RRE donor template is shown (bases 7823-7992 and ΔG = -59.40 kcal/mole). The number 72 marks the major pause site at an A residue.
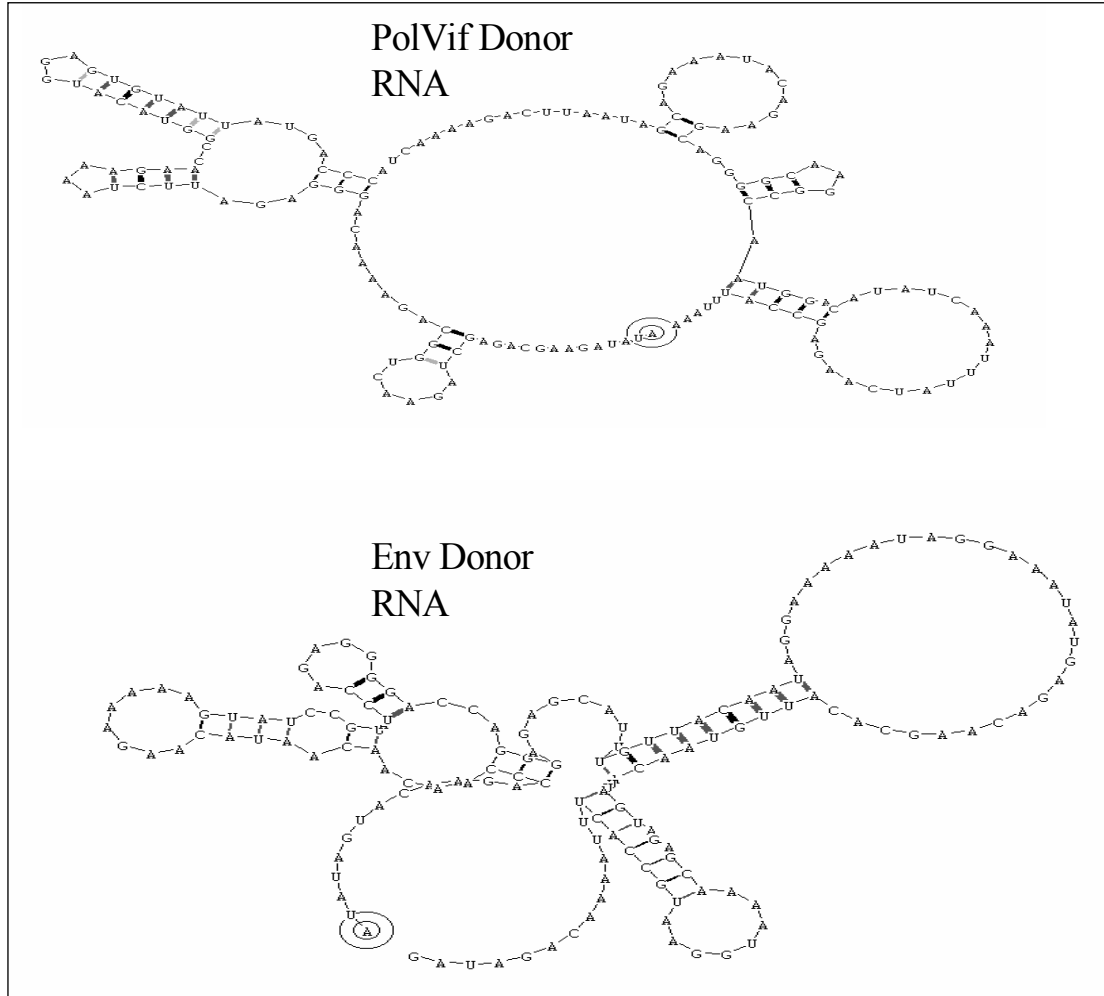
Figure 2-4: Weakly structured templates

Predicted secondary structures of weakly structured donor RNA templates. The base encircled in two rings is the 5' end. The probability of formation for particular base pairs is indicated by the thickness of the lines between them. There are no major pause sites in these two templates. PolVif template (bases 3441-3610 and ΔG =-19.3 kcal/mole) and Env template are shown (bases 7101-7270 and ΔG = -15 kcal/mole)

Effect of NC protein on strand transfer in highly structured RNA substrates:
In this section, autoradiograms of strand transfer assays conducted on three RNA substrates are shown in Figure 2-5, Figure 2-6, and Figure 2-7, and their corresponding graphical quantitations are shown in Figure 2-8. The first pair of donor-acceptor RNA substrates we tested were from the U3 3'LTR region (corresponding to bases 9074-9265 of the HIV-1 provirus as derived from plasmid pNL4-3 [127], see Methods). This region folded to form a complex structure with several strong stem loops and a ΔG= -49.3 kcal/mole as shown in Figure 2-3. A previous *ex vivo* study identified a 24-base segment between bases 9158-9183 in this region as a potential hot spot for recombination [31]. The study used two viral vectors based on different strains of HIV-1 to analyze recombination and obligatory strand transfers in HIV. The vector viruses were harvested from producer cells (CD4-HIV-1 Env inducible cell line) and were used to infect target cells (CD4+ HeLa T4 cells). Of the 86-target cell clones analyzed, 11 clones were shown to undergo homologous recombination during minus strand synthesis in the U3 genome region. DNA sequencing analysis of these clones revealed a 24-base segment as a potential hot spot for recombination. The report also suggested that the region might have RNA secondary structures that induce pausing. Donor and acceptor RNA templates were generated from this zone and used in time course reactions.

Figure 2-5 is an autoradiogram of a PAGE gel showing the resolved DNA products. The band at position 197, as indicated by the molecular ladder on the leftmost lane, are products of strand transfer from the donor to acceptor RNA. The band at position 175 results from completed synthesis on the donor RNAs without any transfer events. Three distinct pause sites P1, P2 and P3 were observed between positions 65 and 90. P1 and P2 are within the 24-base region noted above, while P3 lies just outside at position 65. The major pause sites were mapped using DNA sequencing gels in which sequencing reactions on cDNA were run using the primer from these reactions (data not shown). Interestingly, the characteristics of the position 65 site are strikingly similar to those of another pause site previously characterized [54, 137]. Both sites are located within the stem of a strong stem loop structure. In both cases, RT stalled at a U residue located just behind a series of strong G-C pairs in the stem (refer Figure 2-3). In the absence of acceptor template the paused products persist even at 64 min, especially the 65 base products. Including NC protein without acceptor did not significantly change the profile. In the presence of acceptor template but absence of NC a decrease in the amount of the paused products was observed over time in comparison to reactions without acceptor. An increase in the amount of transfer product was also seen. The results are consistent with paused products transferring to the acceptor template and being extended. Paused products were "chased" more rapidly in the presence of NC and acceptor and the level of transfer product also increased to a greater extent than in reactions with acceptor alone (refer Figure 2-8). Results here are in agreement with earlier reports that have shown that paused DNAs can be focal points of strand transfer [54, 119,

125]. The stimulatory effect of NC on recombination has also been supported by numerous experiments [81, 99].

Transfer efficiency results from the above experiment are graphically shown in Figure 2-8. It is evident from this graph that NC enhances strand transfer significantly in this region of the genome. Each of the experiments was repeated two or more times to confirm the observed trends. Experiments presented are representative of the observations.

Figure 2-5: Time course assay-U3 3'LTR templates

Shown is an autoradiogram of time course assay performed using highly structured U3 3' LTR RNA template. The leftmost lane 'ML' is the molecular ladder indicating the size (in nucleotides) of bands in the other lanes. The reactions were carried out either in the presence or absence of nucleocapsid protein (NC) and acceptor RNA templates (Accp) as indicated above each set of lanes. Donor RNA was used in all reactions. Each reaction was carried out at time points 0, 2, 4, 8, 16, 32 and 64 minutes as shown by the corresponding set of seven lanes from left to right. The transfer and full-length donor-directed DNA products, the major pause sites and the primers are indicated as T, F, P (denoted P1, P2, and P3) and PR respectively.

The second pair of donor-acceptor structured RNAs examined was from the *gag-pol* region (bases 2009-2200, ΔG= -45.30 kcal, Figure 2-3). This region includes the portion of genomic mRNA where the programmed −1 ribosomal frameshifting event used to produce enzymatic proteins (RT, PR, IN) from the pol region occurs. The mechanism allows the virus to maintain a well-regulated ratio of Gag proteins to Gag-Pol proteins in infected cells, which is vital for efficient assembly of infectious particles. The segment of gag-pol region that we used included the heptameric 'slippery site' (UUUUUUA sequence) and a downstream RNA secondary structure, both of which have been shown to promote efficient frameshifting [138, 139]. Earlier literature has suggested that the RNA secondary structure is a simple stem loop [139, 140], but a recent report favors a more complex intramolecular triplex RNA structure [141]. In either case, the structure seems to cause the ribosomes to pause and subsequently slip −1 base over the slippery site. Interestingly, this structure also caused the RT to pause as shown on Figure 2-3. This figure shows a time course reaction conducted on the GagPol substrates. The gel reveals a single strong pause site. Sequencing reactions mapped the site to an A residue at position 77 from the 3' end of the Donor RNA. In this case the A was also part of a strong stem structure and was followed by four G-C pairs. Once again NC significantly stimulated transfer from this region (refer Figure 2-8) and production of transfer products was accompanied by an apparent chasing of the paused DNA product.

Figure 2-6: Time course assay-GagPol templates

Shown is an autoradiogram of time course assay performed using highly structured GagPol template. All markings are the same as the legend of Figure 2-5.

The third donor-acceptor pair of structured RNAs examined was from the RRE region (bases 7801-7992, ΔG= -59.40 kcal, Figure 2-3). The RRE or the Rev Response Element has been characterized by Malim *et al.* [136] as a 234 base region that the Rev protein can bind to. Later on, Mann *et al.* [135] demonstrated that an extra 58 bases at the 5' end and the 59 bases at the 3' end of the original 234 region (a total of 351 bases) was the complete biologically active RRE. The binding of Rev protein to this highly structured region is important for the nuclear export of unspliced and partially spliced HIV mRNAs. A strong pause site was mapped to an A residue at position 72 in this template (refer Figure 2-7 and, Figure 2-3). This site was located just a few bases downstream from the base of a stemloop and GC pairs. In this way it was somewhat different from the pause sites in U3 3'LTR and GagPol templates that were found within stem loops and just before GC pairs.

As with the other structured RNAs examined above, NC enhanced strand transfer in the RRE region that was accompanied by chasing of the major paused DNA products. For all the structured substrates the predicted secondary structures seemed to obstruct RT as indicated by the presence of prominent pause sites. Moreover, the disappearance of pause sites correlated with the presence of acceptor and was further enhanced by NC (refer Figure 2-8). The results suggest that, the pause sites are probably focal points for strand transfers in these particular structured RNAs. Note that this does not imply that all transfers from the RNAs occur from the pause sites, only that transfer is exaggerated from these positions.
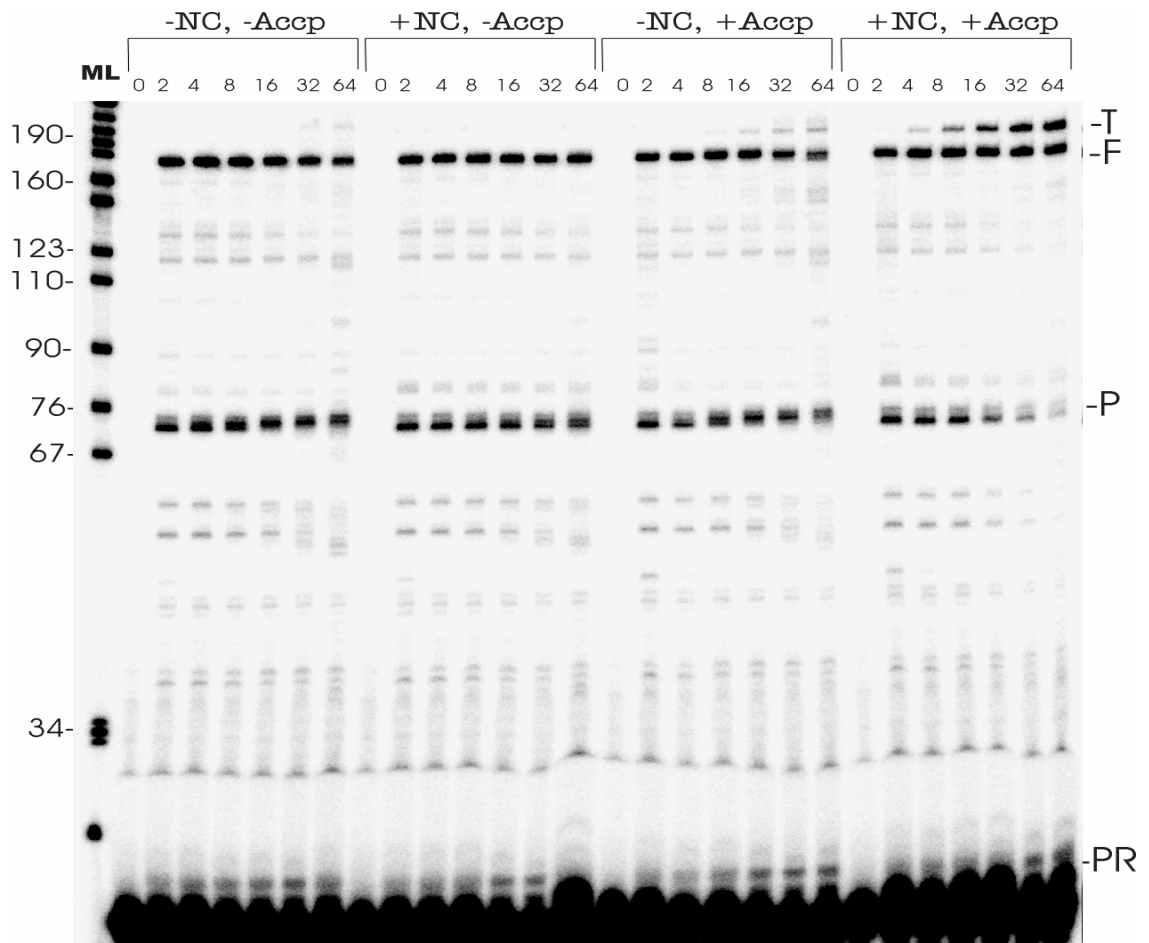
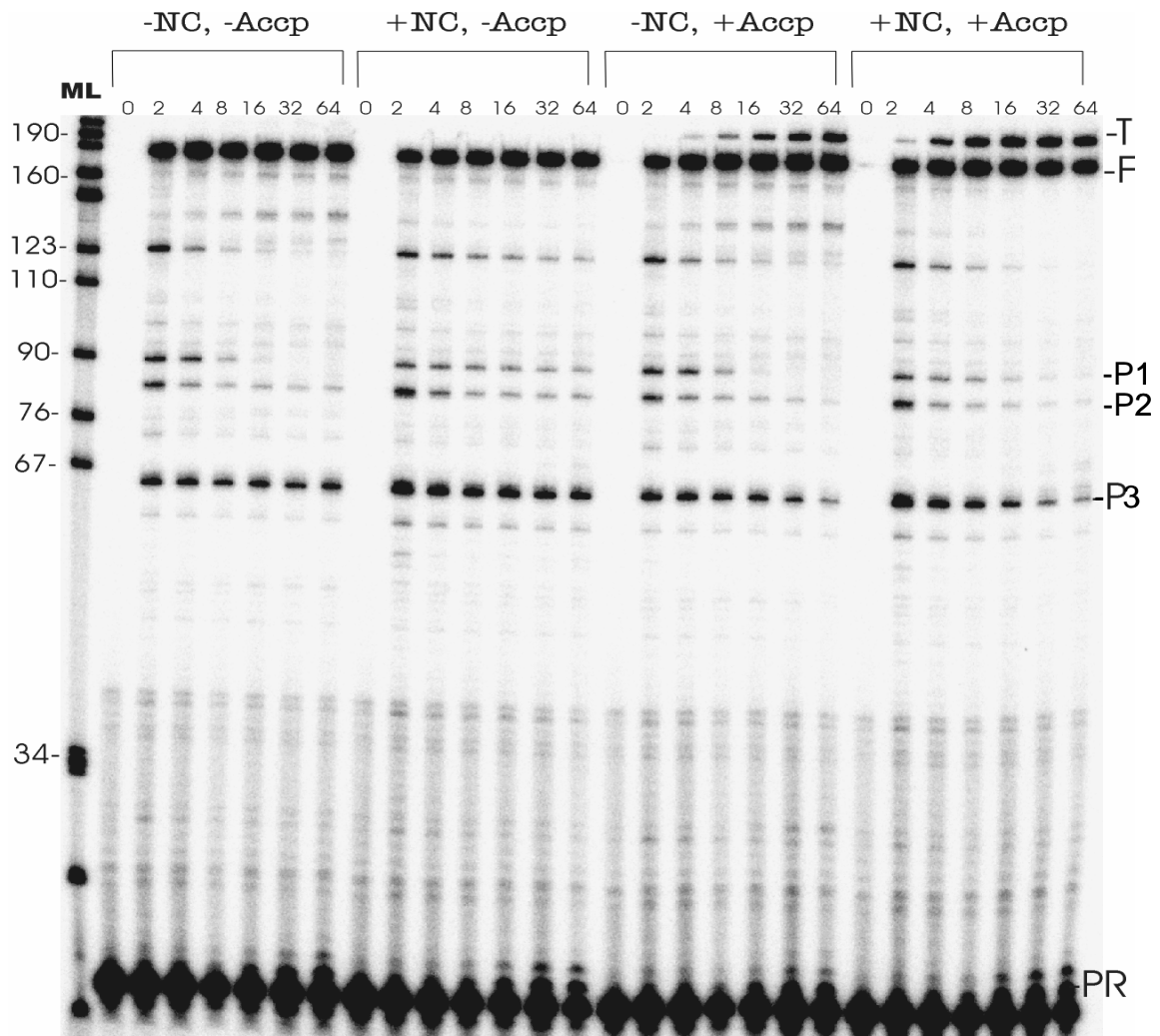**Figure 2-7: Time course assay-RRE templates**

Shown is an autoradiogram of time course assay performed using highly structured RRE template. All markings are the same as the legend of Figure 2-5.

A. U3 3'LTR substrates

B. GagPol substrates
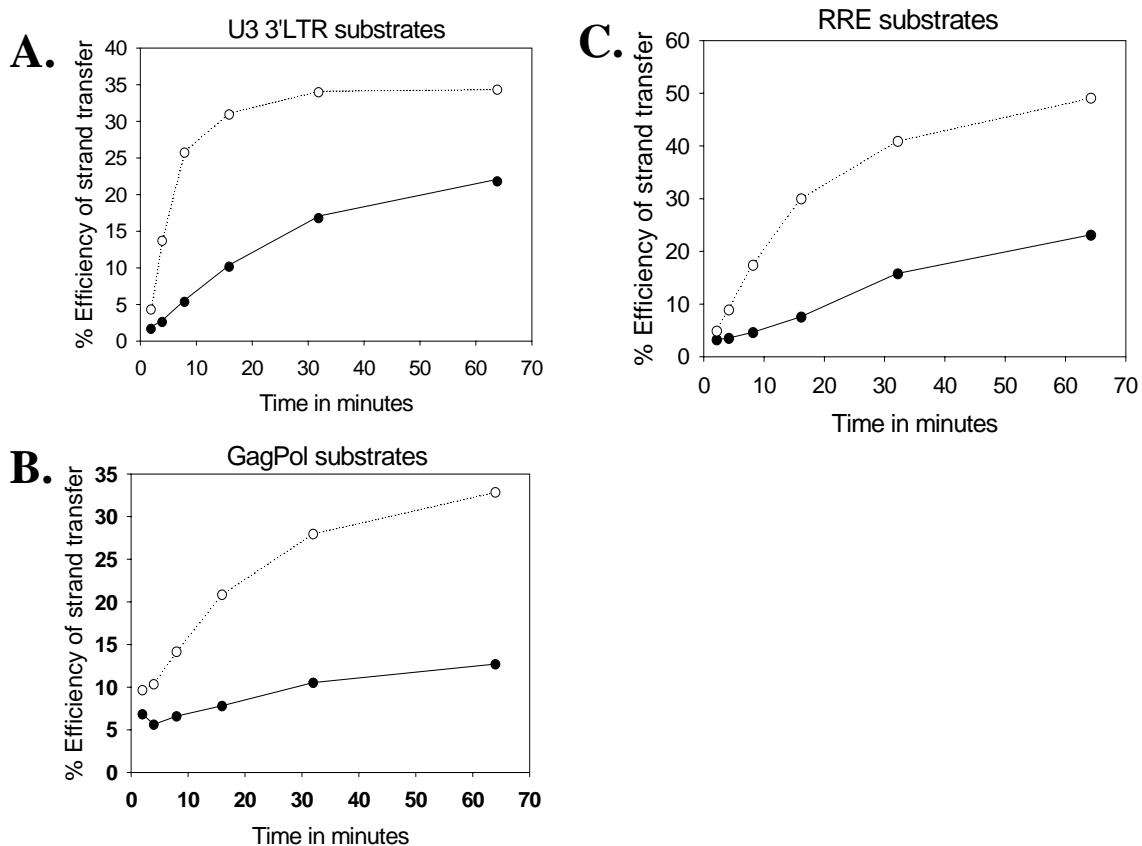
C. RRE substrates

Figure 2-8: Efficiency of transfer-highly structured templates

Graph of efficiency of strand transfer *vs.* time for the highly structured RNA substrates. The percent transfer efficiency was defined as Transfer DNA products (T)/Transfer + Full-length donor-directed products (F), times 100 ((T/(T+F)) x 100). The templates used are indicated above each graph in A, B, and C. The solid circles represent time course reactions conducted in the absence of nucleocapsid protein (-NC) and the presence of acceptor RNA (+Accp). The open circles represent time course reactions conducted in the presence of nucleocapsid protein (+NC) and acceptor RNA (+Accp).

Effect of NC protein on strand transfer in weakly structured RNA substrates -
In this section, autoradiograms of strand transfer assays conducted on two RNA substrates are shown in Figure 2-9 and Figure 2-10 and their corresponding graphical quantitations are shown in Figure 2-11. The first donor-acceptor pair of RNA substrates tested was from the *pol-vif* region of the viral genome (bases 3419-3610, ΔG= -19.3 kcal, Figure 2-4). The experiment shown in Figure 2-9 demonstrates that "strong" pause sites are absent on this donor. Surprisingly, the efficiency of strand transfer on this substrate in the absence of NC was significantly higher than for the more structured substrates described above. In addition, though NC did hasten the onset of strand transfer, the overall enhancement was less as compared to the highly structured RNAs (Figure 2-11).
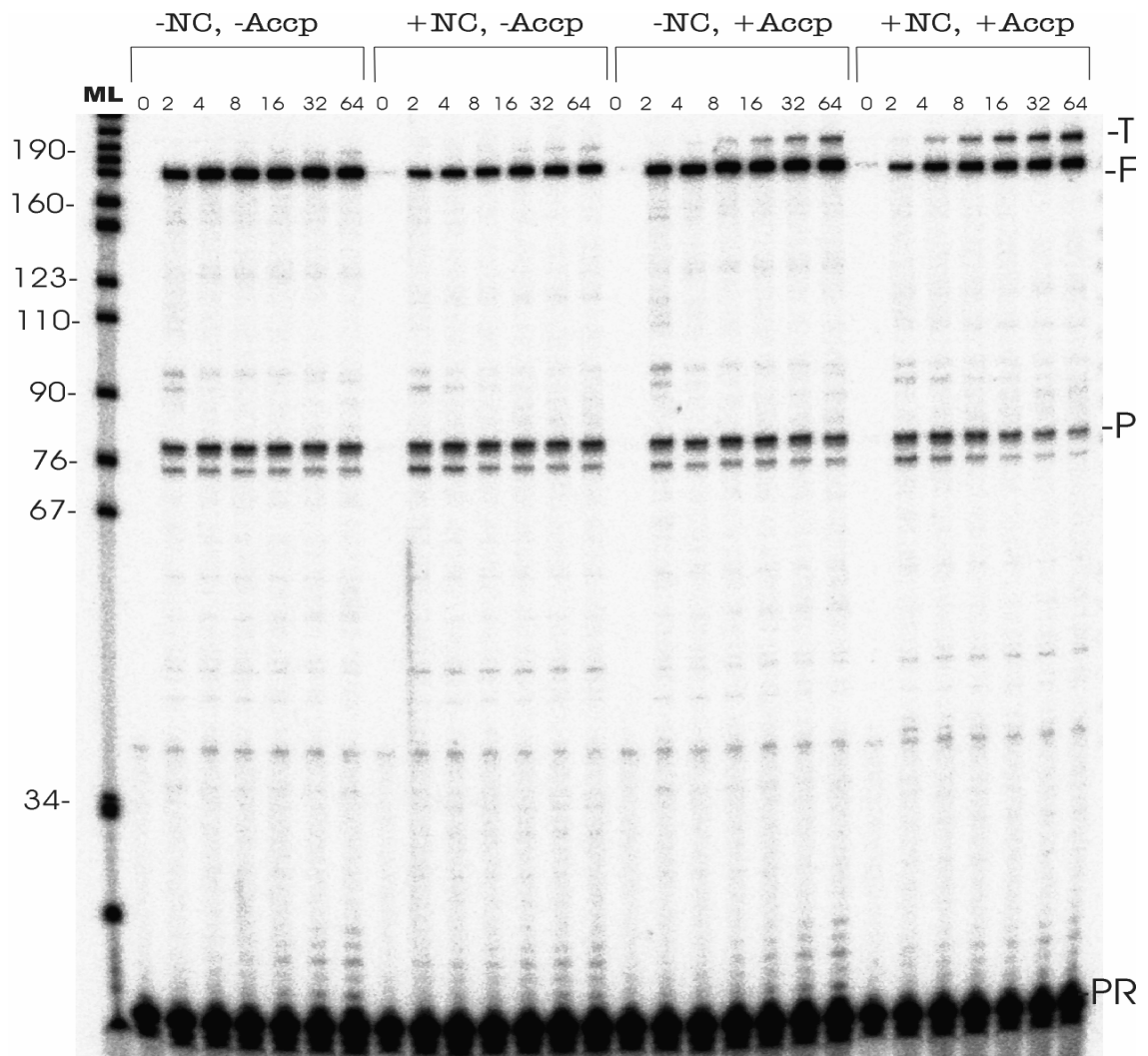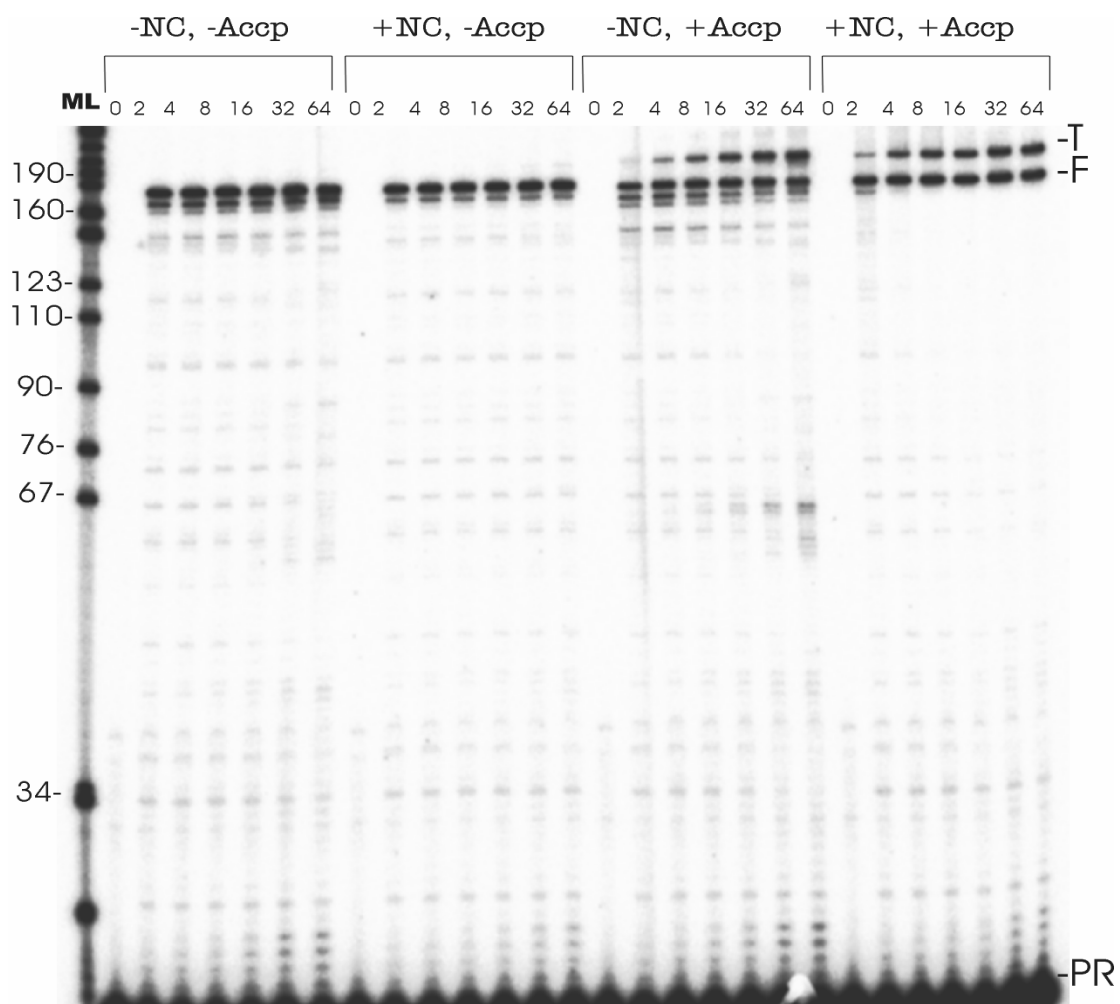
Figure 2-9: Time course assay-PolVif templates

Shown is an autoradiogram of time course assay performed using weakly structured PolVif templates. All markings are the same as the legend of Figure 2-5.

Similar results were obtained with the second low structure donor-acceptor pair of RNA substrates from the *env* region (bases 7079-7270, ΔG= - 15 kcal, Figure 2-4). The region was predicted to have the weakest structure amongst the other four substrates, as shown by the ΔG value. This region also showed an absence of strong pause sites (Figure 2-10). Some weak pause sites were observed, but they quickly disappeared even in the absence of NC and acceptor RNA. Once again, the level of strand transfer was relatively high in the absence of NC and NC moderately increased transfer, but to a lesser extent than on structured substrates as seen in Figure 2-11.
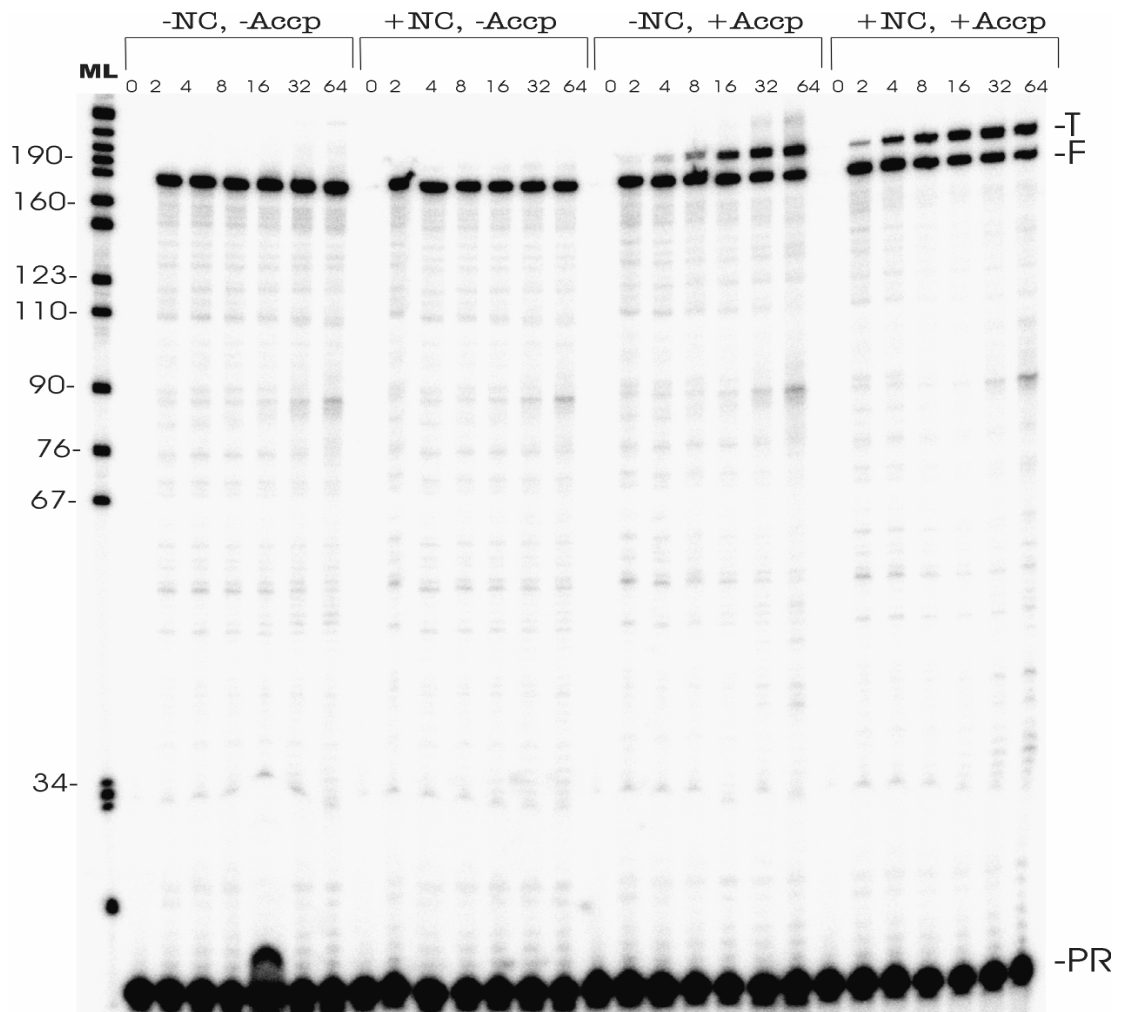
Figure 2-10: Time course assay-Env templates

Shown is an autoradiogram of time course assay performed using weakly structured Env templates. All markings are the same as the legend of Figure 2-5.

A.

B.
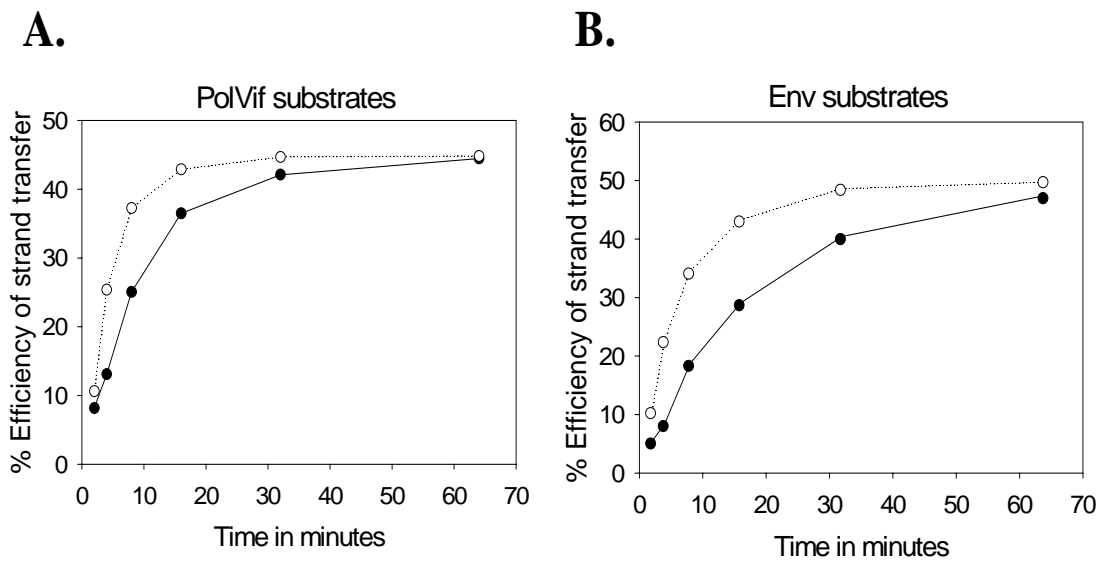
Figure 2-11: Efficiency of transfer-weakly structured templates

Graph of efficiency of strand transfer *vs.* time for the weakly structured RNA substrates. The substrates used are indicated above each graph in A and B. All the other markings are the same as in the legend of Figure 2-8.

Effect of varying concentrations of NC on high or low structure RNA substrates - I conducted strand transfer assays with increasing concentrations of wild type NC protein on three of the substrates; U3 3'LTR, GagPol and PolVif (Figure 2-12). NC concentrations were increased from 0.0625 μM–4 μM. Since 32 min was sufficient to obtain adequate levels of reverse transcription and strand transfer, reactions were allowed to proceed only to this time point. The results are presented graphically in Figure 2-13. Both the U3 3'LTR and GagPol regions showed a significant increase in transfer efficiency as the concentration of NC was increased. The efficiency of transfer approximately doubled with GagPol from 0 to 1 μM NC while U3 3' LTR increased by about 1/3. No further increase was observed beyond 1 μM NC. In contrast, the less structured PolVif region showed a higher level of strand transfer without NC and increased by a relatively small amount when NC was included. Levels of efficiency at 1 μM NC were comparable for PolVif and U3 3' LTR regions while the GagPol region peaked at a lower level. These observations show that NC enhances transfer in a concentration dependent manner and to a greater extent on RNAs with more structure.
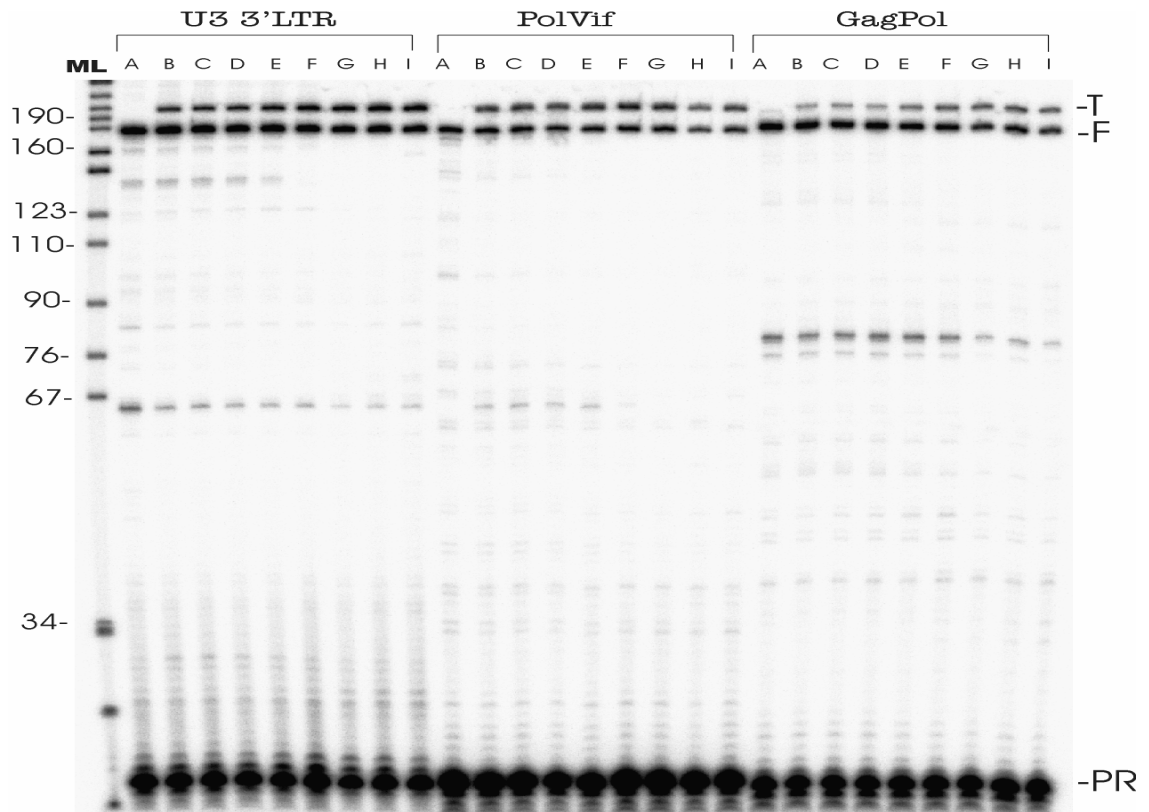
Figure 2-12: NC titration assay

NC titration assay conducted on the U3 3'LTR, PolVif and GagPol RNA substrates. Shown is an autoradiogram of assays that were conducted on three RNA templates as indicated above each set of lanes. ML is the molecular ladder that indicates the size (in nucleotides) of bands in the other lanes. All lanes labeled A are the control lanes with no NC protein and no acceptor RNA. Lanes labeled B contain acceptor RNA but no NC protein. Lanes labeled C to I contain acceptor RNA and NC protein. The final concentrations of NC protein used ranged from 0.0625 μM–4 μM, i.e. lanes C, 0.0625 μM; lanes D, 0.125 μM; lanes E, 0.25 μM; lanes F, 0.5 μM; lanes G, 1 μM; lanes H, 2 μM; and lanes I, 4 μM. The reactions were conducted for 32 minutes. T, F and PR refer to the transfer DNA products, full-length donor-directed DNA products and the primers, respectively
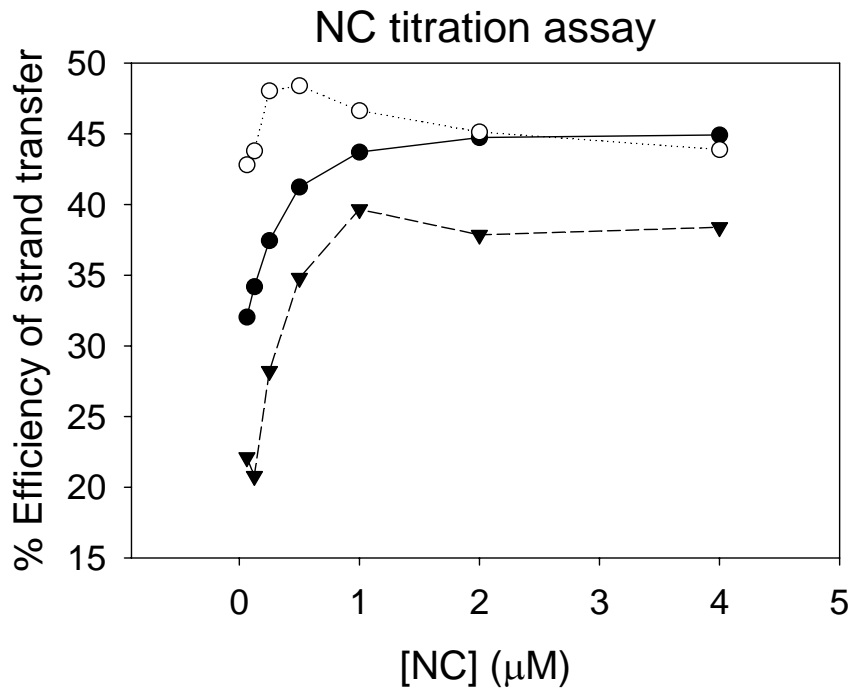
Figure 2-13: Efficiency of transfer-NC titration assay

Shown is a graph of efficiency of strand transfer in U3 3'LTR, PolVif and in GagPol templates *vs.* NC concentration in μM. The percent transfer efficiency was defined as described in Figure 2-8. The solid circles represent assays conducted on U3 3'LTR RNA, the open circles are assays on PolVif RNA, and the solid inverted triangles are assays on GagPol RNA.

2.5 Discussion

In this study an *in vitro* assay that simulates internal strand transfer events occurring during minus strand DNA synthesis was used. The substrates tested transfer over 150-base homologous regions and were chosen to span a wide range with respect to structural strength. This allowed a comparison of transfer on highly structured *vs.* relatively non-structured RNAs in the presence or absence of NC. The level of transfer in the absence of NC was higher on substrates with low structure and without prominent DNA synthesis pause sites. NC only modestly stimulated transfer on these substrates. In contrast, highly structured substrates with prominent pause sites transferred relatively poorly in the absence of NC but were highly stimulated when NC was present. Much of the stimulation on the latter substrates appeared to result from the "chasing" of paused products into transfer products. It was interesting that the low structure substrates showed high levels of strand transfer even though very little pausing was observed.

Both pause-driven transfer [54, 119, 125, 137, 142] and transfer driven by non-pause related mechanism (based on interactions between template structures for example) have been proposed [126, 143, 144]. Chen *et al* [145] have proposed that a pause site can serve as docking sites where the acceptor RNA interacts with the nascent DNA to form a trimeric complex consisting of the acceptor, donor, and DNA. The acceptor then "zips" up the complementary DNA and displaces the donor as RT continues extension of the paused DNA. This leads to transfer at a point beyond the pause site. Alternatively, other reports have proposed that interactions between transferring DNA and specific structures on the acceptor can trigger strand transfers

independent of pausing [126, 143, 144]. For those regions with high structure in the current experiments, at least part of the transfer appeared to result from a standard pause-driven mechanism where DNAs stalled at the pause site transferred to the acceptor. The other regions (PolVif and Env) showed very little pausing, so some other type of mechanism is likely. One possibility is a non-pause-induced mechanism similar to the "zipping" mechanism as proposed by Chen *et al* above. In the above mechanism pausing provides time for the formation of a trimer, which then resolves at a point downstream of the pause site. In low structure regions, pausing may not be necessary for trimer formation because the lack of secondary structures that impede hybridization allows rapid association of the acceptor and nascent DNA. With this in mind, strand transfer could be viewed as a compromise between opposing processes. Pausing favors transfer by providing substrates that can readily transfer to the acceptor. However, the high degree of structure that is generally associated with pausing is not conducive to association between the acceptor and transferring DNA, thus opposing transfer. If most recombination events occurred by non-pause-induced zipping mechanisms, then recombination would likely be fairly homogenously spread over the whole HIV genome and may even focus in regions with low structure. Ultimately, there may be several mechanisms by which strand transfer can occur. The relative quantitative importance of the various mechanisms remains to be determined.

Overall the results showed that strand transfer could occur efficiently by mechanisms that do not involve pausing, although when present, pause sites serve as focal points for transfer events. Results presented here indicate that the role of NC in

recombination is likely dependent on the structure of the region from where strand transfer is occurring. Unwinding activity is especially important in this process because regions with high structure appear to transfer relatively poorly unless NC is present. In low structured regions, there is a relatively less requirement for unwinding activity and this is corroborated by the lowered need for NC protein (refer Figure 2-13). Nevertheless, the annealing component of NC may still be necessary in longer stretches of low structured regions on the genome. These studies suggest a role for structural intricacies of RNA templates in determining the extent of influence of NC on recombination.

Chapter 3   Effect of mutant NC proteins on strand transfer in GagPol and Env RNA templates

3.1 Introduction

All retroviruses NC proteins except those of the spumaretrovirus group contain either one or two copies of a C-X_2-C-X_4-H-X_4-C motif. The X represents variable amino acids. Each motif chelates Zn^{2+} ions through the cysteine (C) sulphurs and the histidine (H) imidazole nitrogen [69, 146] to form structures that are referred to as the first and second zinc finger or N-terminal and C-terminal zinc finger, respectively. The cysteine and histidine residues are very important for specific packaging of viral RNA [61, 147]. Replacement of these residues with others that are unable to chelate zinc result in noninfectious virions that are deficient in packaged viral RNA.

A number of assays have been developed to study the mechanism of nucleic acid chaperone activity of NC protein; one of the methods uses various mutant NC proteins. Gorelick *et al*. [74] have constructed HIV-1 NC zinc finger exchange mutants *via* oligonucleotide-directed mutagenesis. These mutants maintain the native sequence of each zinc finger but their positions in the NC protein were changed. In the NY5/LAV strain or the pNL4-3 clone of HIV-1 NC the zinc fingers differ from each other by five amino acid residues (refer Figure 1-7), hence, they were first sub-cloned into separate plasmids and then mutagenized to exchange the first and second finger nucleotide sequences (designated NC 2.1). In addition to the finger switch mutant, a mutant in which finger one was replaced by finger two (NC with two copies of finger two, designated 2.2) and one with finger two replaced by finger one (NC

1.1) were also constructed. Analyses of the mutants suggest that the two fingers are not functionally equivalent; the first finger was shown to be more important for RNA packaging. Mutant 2.2 was highly deficient in packaging and replication relative to 1.1. Mutant 2.1 was intermediate indicating that the context of finger one is also important. None of the mutants functioned as well as the wild type with respect to replication. This indicated that having one copy of each finger in the proper context is optimal.

A number of other investigators have also tried to determine the exact residues necessary for chaperone activity of NC. Overall, the reports indicate that the basic backbone residues of NC are an absolute requirement for chaperone activity [148, 149]. But many studies done in the context of highly structured RNAs like the TAR stem loop and the tRNA primer [150, 151] have assigned a greater role for first zinc finger; especially in the unwinding component of NC activity. The role of the second zinc finger is still unknown; it may possibly serve a role in the annealing/hybridization component of NC protein. In Chapter 2 it was shown that NC protein had differential effects on strand transfer in highly structured *vs*. weakly structured templates from the viral genome. Low structured templates like the Env require little or no unwinding activity and hence they are ideal substrates for analysis of annealing activity. In this section, strand transfer was conducted on one highly structured and one weakly structured template; the GagPol and Env RNAs, respectively, in the presence of mutant NC proteins 1.1 and 2.2. These experiments were designed to uncover the roles of the two zinc fingers in NC chaperone activity.

3.2 Materials

Plasmid pNL4-3 obtained from the NIH AIDS Research and Reference Reagent Program contains a complete copy of the HIV-1 provirus derived from strains NY5 and LAV [127]. PCR primers were obtained from Integrated DNA Technologies, Inc. Recombinant HIV-RT was graciously provided to us by Genetics Institute (Cambridge, MA) (see Chapter 2 Materials section, page 35 for unit definition). NC finger mutants 1.1 and 2.2 were a gift from Dr. Robert Gorelick (SAIC Frederick, Maryland). These proteins were expressed and purified as described in Carteau et al. [152], and quantified by amino acid analysis on a Beckman Systems 6300 amino acid analyzer (Beckman Coulter, Inc., Fullerton, CA). Wild type NC clone was a gift from Dr Charles McHenry. It was purified and quantified as described in Section 2.3. Taq polymerase was from Eppendorf. SP6 RNA polymerase, DNase I-RNase free, and RNase-DNase free were from Roche Diagnostics. RNase inhibitor was from Promega. T4 Polynucleotide Kinase was obtained from New England Biolabs. Proteinase K was obtained from Kodak. Radiolabeled compounds: γ ^{32}P ATP was obtained from Amersham. Sephadex G-25 spin columns were from Amika Corp. All other chemicals were from Sigma or Fisher Scientific.

3.3 Methods

Preparation of RNA substrates: Donor and acceptor DNAs were first generated from the gag-pol and env regions of the pNL4-3 plasmid with the primers listed in Table 2-1. GagPol and Env donor and acceptor RNAs were prepared from the PCR DNAs by run-off transcription (Section 2.3).

RNA-DNA hybridization: DNA primers that bound specifically to the donor GagPol and Env RNA transcripts were ^{32}P-labeled at the 5' end with T4 polynucleotide kinase according to the manufacturer's protocol. The GagPol donor RNA was hybridized to complementary labeled primer by mixing primer:transcript at approximately 5:1 ratio in 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 80 mM KCl, and 0.1 mM EDTA (pH 8.0). The mixture was heated to 65°C for 5 min and then slowly cooled to room temperature. This donor RNA-primer hybrid (2 nM final concentration of RNA) was preincubated for 3 min with 10 nM of GagPol acceptor RNA template in the presence or absence of wild type NC in 42 µl of buffer at 37°C. Similar reaction mixtures were preincubated in the presence of 1.1 NC or 2.2 NC (as indicated). One molecule of NC per two nucleotides was used in the reactions, i.e. final concentrations of 1.16 µM and 0.27 µM of wild type NC and each mutant NC. The reactions were initiated by the addition of 8 µl of HIV-RT at a final concentration of 2.5-units/µl. The time course experiments were conducted as previously described (Section 2.3). The experiments were repeated using the Env templates.

3.4 Results

Two different mutants of NC protein were tested to determine their effect on transfer from structured (GagPol) and non-structured (Env) regions. Shown in Figure 3-1 are strand transfer experiments with wild type or mutants 1.1 or 2.2 NC, performed on the Env substrates. Plots from this experiment are shown in Figure 3-3; the percent transfer efficiency was defined as Transfer DNA products (T)/Transfer + Full-length donor-directed products (F), times 100 ((T/ (T+F)) x 100). Wild type and 1.1 showed comparable levels of transfer on the low structure Env substrate while 2.2 enhanced transfer somewhat better than the others. This was in contrast to strand transfer on the highly structured GagPol substrate Figure 3-2 and Figure 3-3. In this case, 2.2 was the least stimulatory while 1.1 and wild type were comparable. The results are consistent with 2.2 retaining annealing activity but having less unwinding activity than 1.1 or wild type. This supports a role for finger 1 in unwinding. The role of finger 2 in annealing is less clear from these experiments since the mutant without finger 2 (1.1) was nearly as good as wild type on both substrates.

3.5 Discussion

Results of annealing assays published by Heath *et al*. [153] suggest that the two zinc fingers may have disproportionate roles in annealing and unwinding with finger one possessing most of the unwinding activity and finger two serving a supportive role. Results reported above showed that mutant 1.1 (two copies of finger 1) enhanced strand transfer to about the same level as wild type on both the highly structured (GagPol) and low structured (Env) substrates (Figure 3-1 and Figure 3-2). In contrast 2.2 (two copies of finger 2) showed greater enhancement than wild type on the Env substrate while stimulation was reduced on GagPol (refer Figure 3-3). These results are consistent with a role for finger 1 in unwinding as the loss of this finger resulted in reduced transfer in the more structured substrate. The results are at least partly consistent with a previous report using these mutants [150]. In that study the mutants were evaluated using a system to test strand transfer of the TAR containing minus strand strong-stop DNA (-ssDNA). In those experiments mutant 1.1 retained 50% of wild type activity and 2.2 showed a complete loss of activity. The differences observed between these two sets of experiments could have resulted from the substrates used. The –ssDNA substrate forms a very strong complex secondary structure which may require a high degree of unwinding and annealing activity to overcome the strong structure. Also, –ssDNA has a tendency to self-prime, an effect that prevents transfer but is inhibited by NC [102, 154]. Therefore this substrate may be very sensitive to small changes in NC that have an effect on self-priming and/or annealing and unwinding. Overall, however, both sets of experiments support a role for finger 1 in helix-destabilization.
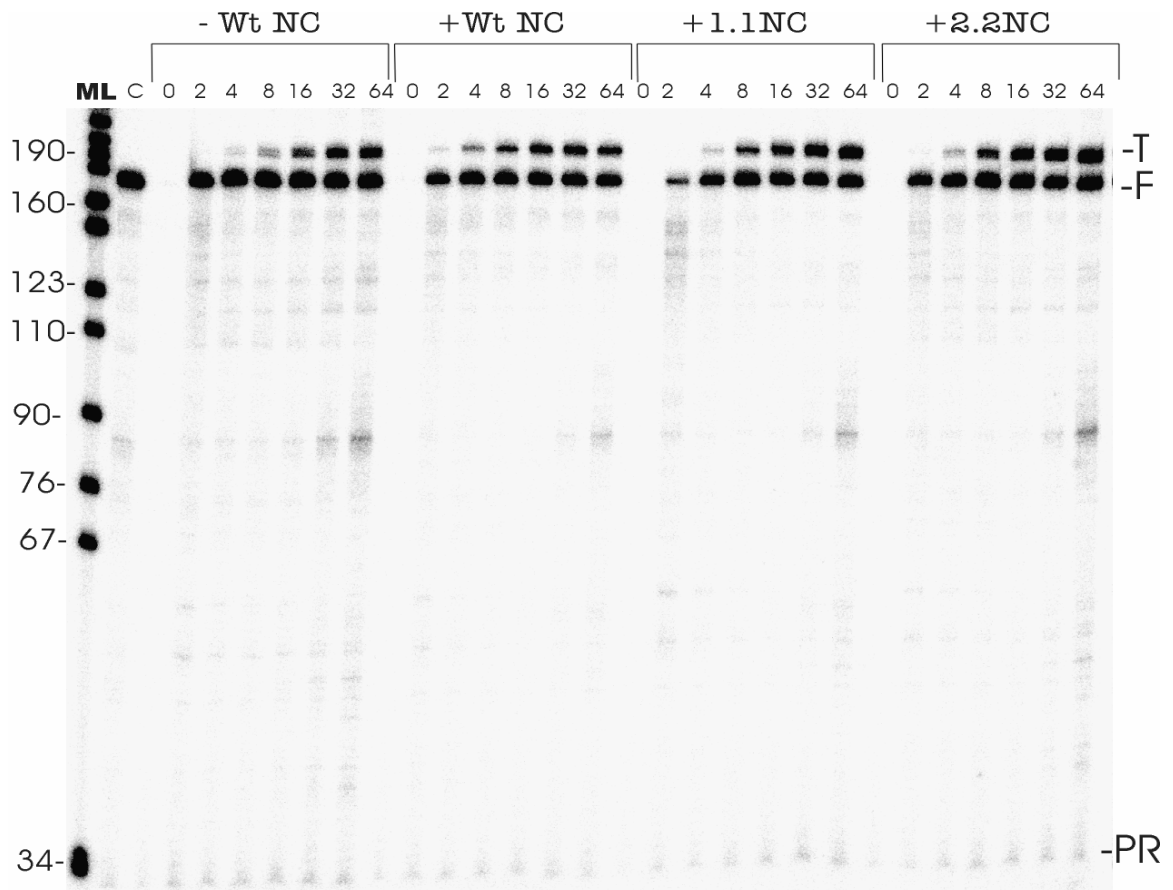
Figure 3-1: NC mutants-Env templates

Shown is the autoradiogram of an assay that was conducted using standard conditions on the Env RNA substrate. Assay was conducted in the absence or in the presence of wild type NC (Wt NC), mutant 1.1, or mutant 2.2 NC as indicated above each set of lanes. Lanes labeled C are controls with no acceptor RNA or NC protein. All the other markings are the same as in legend of Figure 2-5.
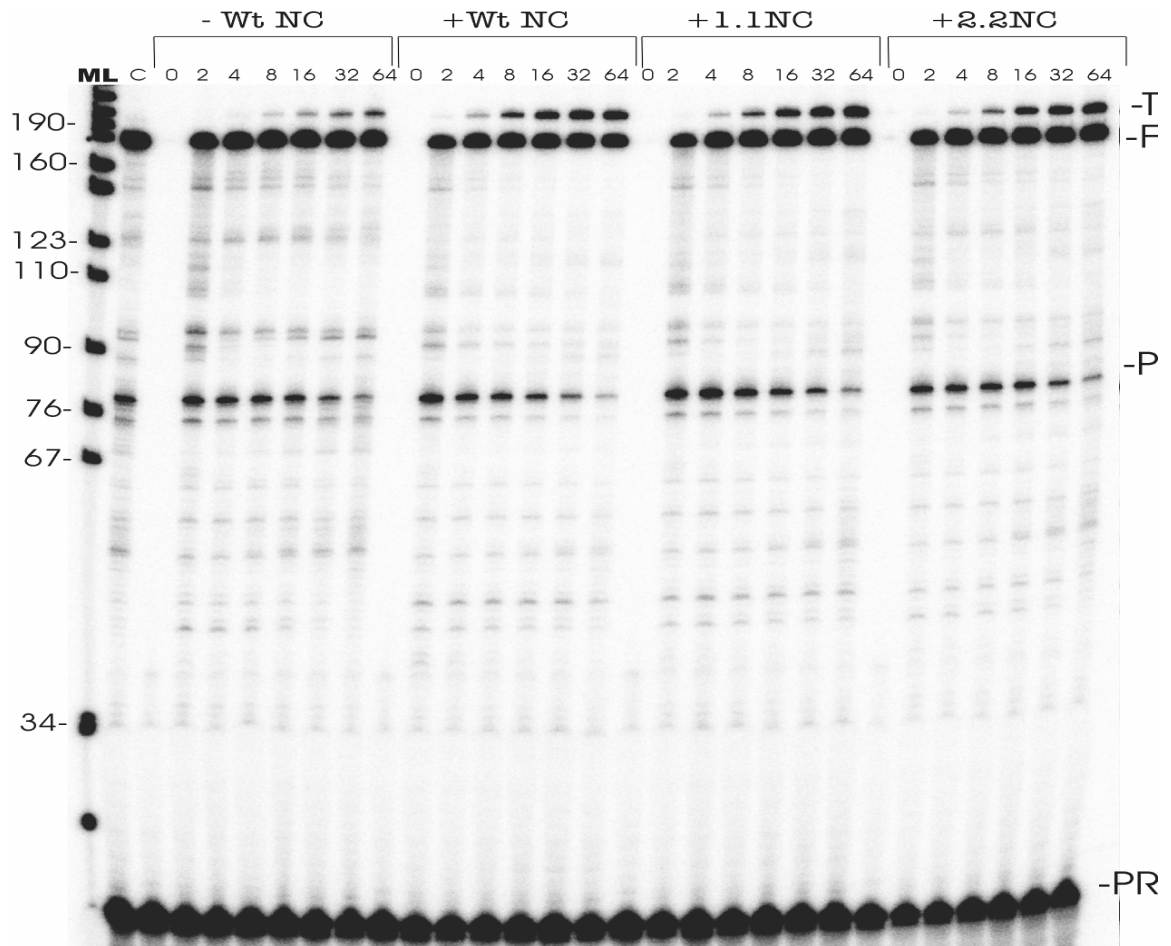
Figure 3-2: NC mutants-GagPol templates

Shown is the autoradiogram of an assay that was conducted using standard conditions on the GagPol RNA substrate. Assay was conducted in the absence or in the presence of wild type NC (Wt NC), mutant 1.1, or mutant 2.2 NC as indicated above each set of lanes. Lanes labeled C are controls with no acceptor RNA or NC protein. All the other markings are the same as in legend of Figure 2-5**.**

A.

Env with Wild type, 1.1 and 2.2 NC

B.

GagPol with Wild type, 1.1 and 2.2 NC

- ●— -Wild type NC
- ○···· +Wild type NC
- ▼--- +1.1 NC
- ▽--- +2.2 NC
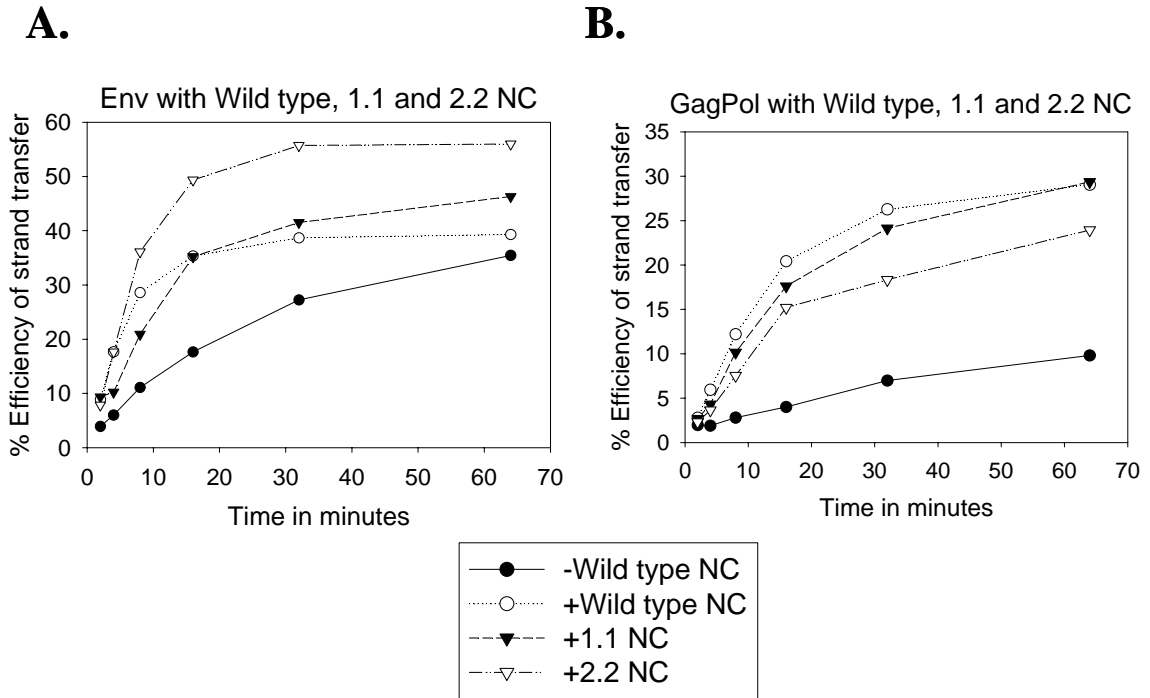
Figure 3-3: NC mutants-Efficiency of transfer

The percent transfer efficiency was defined as described in Figure 2-8. Results were derived for the experiments shown in Figure 3-1 and Figure 3-2. The substrates used are indicated above each graph in A and B.

Chapter 4   Effect of acceptor RNA structure on 'pausing' and strand transfer

4.1   Introduction

The current data garnered from all the previous analyses (Chapters 2 and 3) leads to the basic hypothesis that pause sites can serve as the focal point for strand transfer, but low structured regions devoid of pause sites can promote transfer by rapid association of DNA and the acceptor RNA. Consequently, NC protein that is known to unwind RNA structures [106] shows modest stimulation of transfer in low structured regions, whereas, there is high level of stimulation for structured regions. If the hypothesis is correct then decreasing the folding strength of a highly structured acceptor like GagPol should enhance the level of strand transfer. Some investigators [126] have proposed that the acceptor RNA structure has a pronounced effect on strand transfer. The goal of the following experiments is to verify whether the acceptor RNA structure has a positive or negative influence on recombination. The GagPol region was chosen for several reasons; (1) it is highly structured with a $\Delta G =$ -35.33 kcal/mole; (2) it has a major pause site that is 'chased' into transfer products in the presence of acceptor and NC; (3) it is a vital ribosomal frameshift region for which information on structure and function are readily available [138, 155].

When the GagPol acceptor RNA structure was compared against the donor RNA the following results were obtained; (i) structure prediction tools like RNAdraw and Zuker's RNA folding programs [130-132] (used in Chapter 2) revealed a strong base pairing of the six Cs and one G between bases 149-158 at the 3' end with complementary bases at positions 19-28 at the 5' end of the GagPol acceptor RNA

(Figure 4-4); this structure was absent in the GagPol donor RNA, (ii) RNase mapping of the GagPol acceptor template showed a protection of the RNA at the above mentioned bases, and (iii) this 5'-3' interaction created an extra stem structure outside the homologous regions of the GagPol acceptor RNA that was absent on the GagPol donor RNA. This was not the case for the other four donor-acceptor pairs that were studied in Chapter 2. Since an interaction of this type would lead to more stable folding, it could potentially interfere with transfer of DNA from the donor to acceptor.

In order to synchronize the GagPol donor-acceptor structures, 25 bases at the 5' end of the acceptor were replaced with sequences that eliminated the 5'-3' interaction. This replacement did not affect the region of homology between the donor and acceptor templates. The new modified acceptor GagPol RNA structure was confirmed by structure analysis tools and RNase mapping (Figure 4-4). This structure, henceforth referred to as 5' modified GagPol acceptor RNA, closely resembled the donor and was used instead of the original acceptor in all experiments involving GagPol substrates.

In this section, the 5'modified GagPol acceptor and four other structures with truncations/mutations involving 5' and 3' ends were used to assess the overall effect of acceptor structure on 'pausing' and strand transfer (Figure 4-1). Overall, the results supported the hypothesis that less structured acceptors allowed a greater level of strand transfer.

4.2 Materials

All the primers that were used to make the different GagPol acceptor templates are listed in Table 4-1. The primers were obtained from Integrated DNA Technologies, Inc. Plasmid pNL4-3, obtained from the NIH AIDS Research and Reference Reagent Program contains a complete copy of the HIV-1 provirus derived from strains NY5 and LAV [127]. Taq polymerase was from Eppendorf. SP6 RNA polymerase, DNase I-RNase free, and RNase-DNase free were from Roche Diagnostics. RNase inhibitor was from Promega. T4 Polynucleotide Kinase was obtained from New England Biolabs. Proteinase K was obtained from Kodak. Radiolabeled compounds: γ ^{32}P ATP was obtained from Amersham. Sephadex G-25 spin columns were from Amika Corp. Recombinant HIV-RT was provided to us by Genetics Institute (Cambridge, MA) (enzyme unit definition is as given in Section 2.2). Aliquots of HIV-RT were stored frozen at -70°C, and a fresh aliquot was used for each experiment. HIV-NC clone was a generous gift from Dr. Charles McHenry (University of Colorado). NC was purified and quantified as explained in Section 2.2. Aliquots of NC were stored frozen at -70°C, and a fresh aliquot was used for each experiment. T1 RNase and RNase A enzymes were obtained from Roche. All other chemicals were from Sigma or Fisher Scientific.

4.3 Methods

PCR amplification of substrates: The GagPol donor was generated using the primers; 5'-**gatttaggtgacactatag***tcaaa*gaaatgtggaaagga-3' and 5' ttgttgtctctaccccagac 3' amplifying from bases 2031-2200 of the HIV-1 provirus as derived from the pNL4-3 plasmid. The bases in bold are the sequences of SP6 promoter and those in italics are non-HIV derived bases that were added to the end of the donor to prevent DNAs synthesized to the end of the donor from transferring to the acceptor. Five different GagPol acceptor RNAs of varying lengths were generated using the PCR primers listed in Table 4-1 (Figure 4-1, Figure 4-2, and Figure 4-3). An SP6 promoter sequence was included on some of the primers to allow transcription of the DNA by SP6 RNA polymerase. PCR reactions were performed according to the enzyme manufacturer's protocol using the provided buffer. One hundred pmoles of each primer was used. The cycling parameters and the protocol for DNA purification are the same as explained in Section 2.3.

Preparation of RNA substrates: Run-off transcription (performed according to the enzyme manufacturer's protocol) was conducted using approximately 5 μg of the purified PCR DNAs and SP6 RNA polymerase enzyme to generate donor RNA transcripts of 175 nucleotides. Five acceptor RNA transcripts of 175, 93, 93, 74 and 55 nucleotides respectively were generated in the same way. The transcription reactions were purified and the RNAs recovered according to protocol explained in Section 2.3.

RNA-DNA hybridization: DNA primers that bound specifically to the donor RNA transcripts were ^{32}P-labeled at the 5' end with T4 polynucleotide kinase

according to the manufacturer's protocol. The donor RNA was hybridized to a complementary labeled primer by mixing primer:transcript at approximately 5:1 ratio in 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 80 mM KCl, and 0.1 mM EDTA (pH 8.0). The mixture was heated to 65°C for 5 min and then slowly cooled to room temperature. Donor RNA-primer DNA hybrid mixtures were prepared for each of the five acceptor RNAs. Time course reactions and gel electrophoreses were performed for each set of GagPol donor-acceptor templates as described in Section 2.3.

RNase protection assays: The structural features of the original GagPol, 5' modified, 5' truncated and 5' truncated mutant GagPol acceptor RNAs were tested by RNase protection assays. The two enzymes used were T1 RNase and RNase A. The recovered RNAs from the above transcription reactions were dephosphorylated using calf-intestinal alkaline phosphatase. The alkaline phosphatase was heat inactivated at 65°C for 10 minutes. Reactions were extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with ethanol. The RNA precipitates were labeled at the 5' end with ^{32}P using T4 polynucleotide kinase. The labeled RNAs were again gel-purified on 8% denaturing gels. The recovered labeled RNAs were quantified spectrophotometrically from optical density.

T1 RNase enzyme analysis was conducted by digesting each labeled RNA from above in a 10 μl reaction mixture containing final concentrations of Tris-HCl (pH 8.0) at 100mM, RNA at 0.2 pmol, serially diluted T1 enzyme and DEPC water. T1 enzyme was diluted to 500 milliunits, 1, 5, 10 and 15 units with a buffer containing final concentrations of 300 mM sodium chloride, 10 mM Tris (pH 7.6), and 5 mM EDTA. The reactions were incubated at room temperature for four

minutes; they were stopped with 5 µl of 2X formamide dye (90% formamide, 10 mM EDTA (pH 8.0), 0.1% xylene cyanol, 0.1% bromophenol blue) and put on ice until they were gel electrophoresed. The samples were resolved on an 8% denaturing polyacrylamide gel containing 7 M urea. Base ladders, G-ladders and undigested control RNA were run along with the samples. The gels were dried and subjected to autoradiography.

The RNase A enzyme analysis was conducted only on the original and 5' modified GagPol acceptor RNA by digesting the labeled RNA in a 10 µl reaction mixture containing final concentrations of sodium acetate (pH 5.2) at 50 mM, RNA at 0.2 pmol, diluted RNase A enzyme and DEPC water. RNase A was diluted to 0.1, 025, 0.5 and 1 and 3 µunits with a buffer containing 1 M sodium acetate (pH 5.2). The reactions were incubated at room temperature for four minutes; they were stopped with 5 µl of 2X formamide dye (90% formamide, 10 mM EDTA (pH 8.0), 0.1% xylene cyanol, 0.1% bromophenol blue) and put on ice until they were gel electrophoresed. The samples were resolved on an 8% denaturing polyacrylamide gel containing 7 M urea. Base ladders, G-ladders and undigested control RNA were run along with the samples. The gels were dried and subjected to autoradiography.

Base ladders for each RNA were prepared by digesting 1 µl of 0.2 pmol/µl of labeled RNA suspended in 3 µl of RNase free water with 1 µl of 500 mM sodium hydroxide in a 65°C heat block for 15 seconds. The reaction was stopped by adding 1 µl of 500 mM HCl. A 15 µl solution containing 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 80 mM KCl, and 0.1 mM EDTA (pH 8.0).was added to this mixture followed by 20 µl of 2X formamide dye (90% formamide, 10 mM EDTA (pH 8.0), 0.1%

xylene cyanol, 0.1% bromophenol blue). The reaction was repeated to generate a ladder for a one minute digestion.

G ladders were generated by digesting 1 µl of 0.2 pmol/µl of labeled RNA with 0.5, 1 and 5 units of T1 enzyme each for 10 minutes at 37°C in a total volume of 5 µl containing final concentrations of 300 mM NaCl, 10 mM Tris-HCl (pH 7.5) and 5 mM EDTA. A 15 µl solution containing 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 80 mM KCl, and 0.1 mM EDTA (pH 8.0).was added to this mixture followed by 20 µl of 2X formamide dye (90% formamide, 10 mM EDTA (pH 8.0), 0.1% xylene cyanol, 0.1% bromophenol blue).

The base ladders and the G ladders prepared above were run along with T1 RNase and RNase A enzyme analyses products on 8% denaturing polyacrylamide gel containing 7 M urea.

Table 4-1: List of five variable GagPol acceptor templates

RNA templates marked A-F[a]	Total length (nt)[b]	Region of homology (nt)[c]	Transfer DNA (nt)	Sequences of PCR primers in 5' to 3' orientation
A. 5'modified GagPol acceptor (2031-2180)	175	150	195	5'**gatttaggtgacactatag** *tattattggatatatatatgcgcgc*gaaatgtggaaaggaagg 3' 5'ctgaagctctcttctggtgg 3'
B. 5'truncated GagPol acceptor (2093-2180)	93	88	113	5'**gatttaggtgacactatag** *tatat*ggaagatctggccttcc 3' 5'ctgaagctctcttctggtgg 3'
C. 5'truncated mutant GagPol acceptor (2111-2180)	93	70	113	5'**gatttaggtgacactatag** *ttattggatatatatatgcgcgc* acaagggaaggccagggaattttc 3' 5' ctgaagctctcttctggtgg 3'
D. 3'-19 GagPol acceptor (2111-2161)	74	51	113	5' **gatttaggtgacactatag** *ttattggatatatatatgcgcgc* acaagggaaggccagggaattttc 3' 5'gggctgttggctctggtctg 3'
E. 3'-38 GagPol acceptor (2111-2142)	55	32	113	5' **gatttaggtgacactatag** *ttattggatatatatatgcgcgc* acaagggaaggccagggaattttc 3' 5'gctctgaagaaaattccctg 3'

(nt) – In nucleotides

a- Refers to sequence numbering of plasmid pNL4-3 that contains an almost complete copy of the HIV-1 provirus derived from strains NY5 and LAV.

b-Total length = length of region of homology + length of non-viral sequences, where region of homology corresponds to viral sequences amplified from pNL4-3 plasmid and non-viral sequences correspond to the italicized sequences on the primers. For example, in template A, total length of 175 = 150 bases from 2031-2180 on pNL4-3 + 25 non-viral bases.

c– Length of region of homology between donor template and the corresponding acceptor template. This is also the length of the HIV sequences that were amplified from pNL4-3 plasmid.

GagPol Donor (175 nt)

3' | 20 | 150 | 5 | 5'

A. 5' modified GagPol acceptor (175 nt)

3' | 150 | 25 | 5'

B. 5' truncated GagPol acceptor (93 nt)

3' | 88 | 5 | 5'

C. 5' truncated mutant GagPol acceptor (93 nt)

3' | 70 | 23 | 5'

D. 3'-19 GagPol acceptor (74 nt)

3' | 51 | 23 | 5'

E. 3'-38 GagPol acceptor (55 nt)

3' | 32 | 23 | 5'
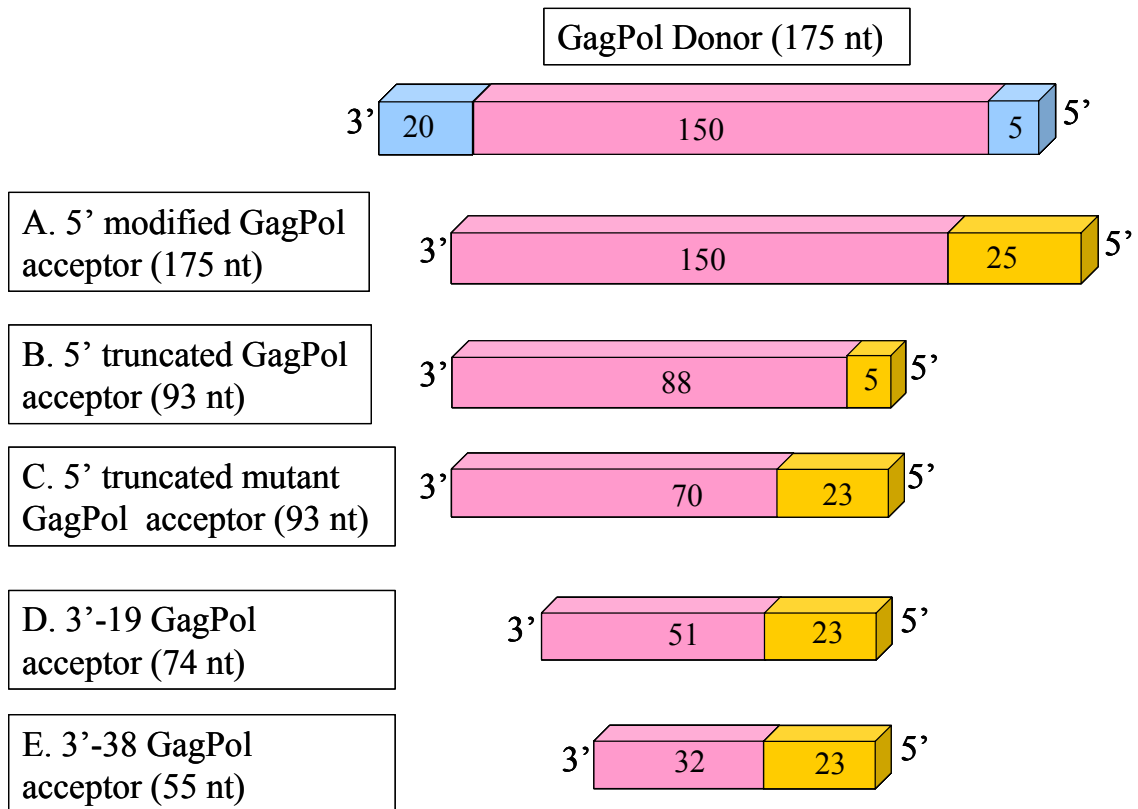
Figure 4-1: GagPol acceptor templates

Shown above is a list of the various acceptor templates (A-E) that were individually tested with the GagPol donor (topmost) in strand transfer assays. The pink areas are the regions of homology between the donor and the corresponding acceptor template. The non-homologous regions are shown in blue on the donor and yellow on the acceptor templates, respectively. The numbers within each colored area is the length in nucleotides (nt).
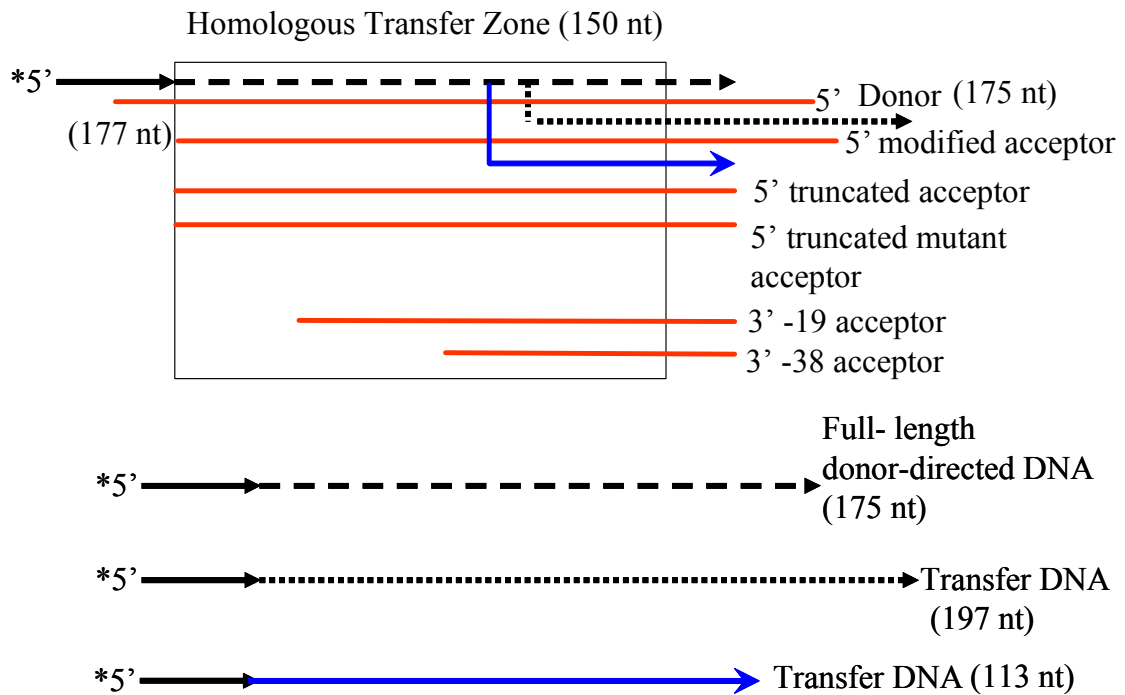
Figure 4-2: Transfer products with various GagPol acceptor RNAs

Shown above is a diagram of strand transfer assay with various GagPol acceptor RNAs as indicated. The area within the box is the region of homology between the GagPol donor and the different acceptor RNAs. The thick line with asterisk at the 5' end represents the radio labeled primer that begins DNA synthesis. The lengths of the full-length donor-directed and transfer DNA are shown in nucleotides (nt).
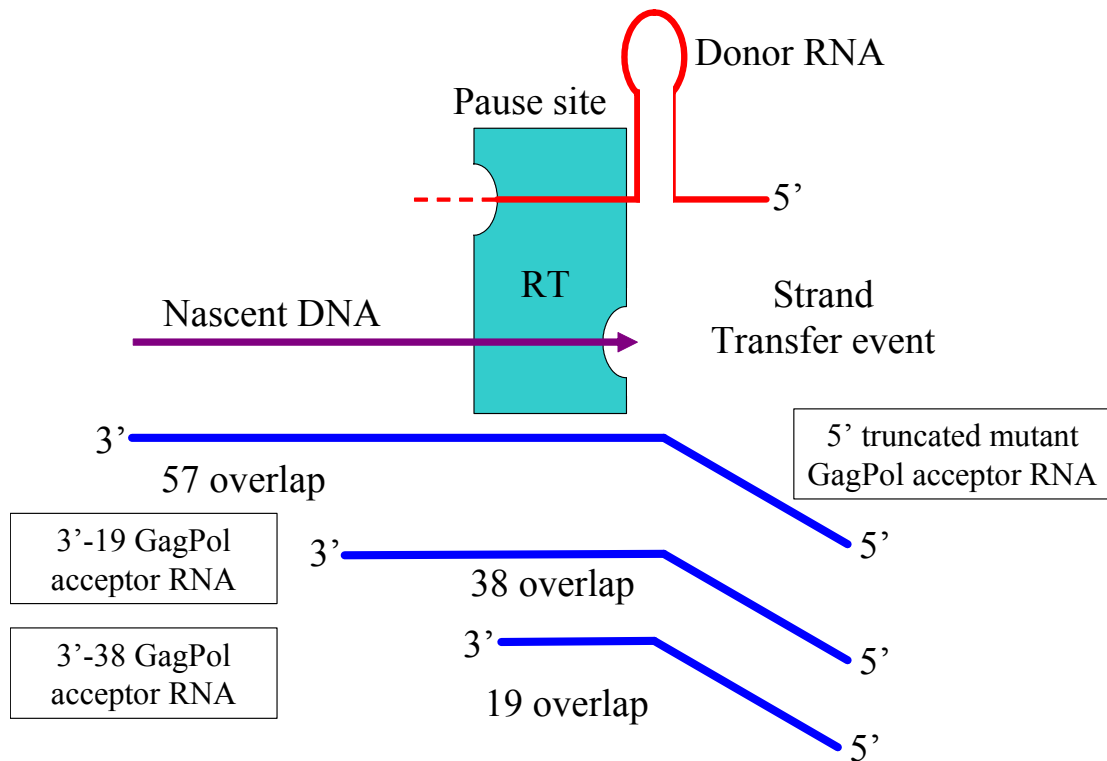
Figure 4-3: Lengths of overlap between GagPol donor and acceptor RNAs

The donor RNA is shown as a red line. Broken red lines are cleavage products of RNase H activity. RT is reverse transcriptase enzyme. Different acceptor RNAs (blue lines) are aligned with the nascent DNA (dark purple line). Each acceptor is capable of forming a hybrid equal to length of overlap (57, 38 or 19) with the nascent DNA before arriving at the pause site. These acceptor RNAs were used to test for acceptor invasion model of transfer in GagPol templates.

4.4 Results

Construction of GagPol donor and acceptor templates: All the PCR primers that were used to generate five variable acceptor templates are listed in Table 4-1. The sequences in bold are the SP6 promoter sequences that allow transcription of the DNA by SP6 RNA polymerase. All the sequences in italics (varying in length from 5-25 nucleotides) are non-HIV derived nucleotides that were introduced to create a non-homologous region at the 5' end of the donor and acceptor. This prevents strand transfers from the end of donor and restricts transfers to internal regions. On primers for the 5' truncated acceptor, the italicized sequences are 5 in number, on the 5' modified acceptor primers they are 25 in number and on all the other primers they are 23 in number. The longer sequences not only introduced non-homologous regions on the 5' end of their templates but served other roles as well; on 5' modified template they aided in making the acceptor and donor look structurally alike (for the reasons stated in the Introduction) and on the 5' truncated mutant, the 3' -19 and the 3' -38 templates their purpose was to disrupt the stem loop structure that carries the major 'pause site' (see below).

The GagPol donor template (ΔG= -35.33 kcal/mole) is 175 bases long; it is made up of 170 bases from positions 2031-2200 of *gag-pol* region of the viral genome and the extra five non-HIV bases at the 5' end. The folding strength was calculated for only 155 bases of the donor template. The remaining 20 bases would be involved in RNA-DNA hybridization with radio labeled primers during synthesis; they would not contribute to the folding of the rest of the template. This template was

paired up with each of the five acceptor templates to test the effect of acceptor structure on strand transfer and 'pausing' in this region of the genome. In each scenario, the composition of the acceptor template determined the length of the homologous region between the donor-acceptor pairs and the transfer DNAs obtained (Figure 4-1). The lengths of transfer products that are obtained from these different acceptors are shown in Figure 4-2. The lengths of overlap between some of these acceptors and the nascent DNA before arriving at the pause site are shown in Figure 4-3. The 175 nucleotides long 5' modified GagPol acceptor (ΔG= -37.22 kcal/mole) is made up of 150 bases from positions 2031-2180 of the genome and an extra 25 non-HIV bases at the 5' end (Figure 4-4). This template was specially designed to be structurally similar to the donor template. The 93 nucleotides long 5' truncated GagPol acceptor carries 88 bases from positions 2093-2180 of the genome and an extra 5 non-HIV bases at the 5' end. This structure has a ΔG value of -29 kcal; it was designed to retain the strong stem loop structure that bears the pause site. The region of homology was shortened from 150 to 88 at the 5'end of this template. The 93 nucleotide long 5' truncated and mutated GagPol acceptor carries 70 bases from positions 2111-2180 of the genome and an extra 23 non-HIV bases at the 5' end. This structure has a ΔG value of -9.5 kcal; it was designed to eliminate the strong stem loop structure while retaining the pause site. The region of homology was shortened to 70 bases at the 5' end (Figure 4-5). The 3' -19 and 3' -38 acceptors are essentially the same as the 5' truncated mutant with additional truncations of 19 and 38 bases at the 3' end and ΔG values of -5.69 kcal/mole and -2.58 kcal/mole

89

respectively. The region of homology is reduced to 51 and 32 bases at the 3' end of the templates, respectively (Figure 4-6).

RNase protection assay of the original GagPol and the 5'modified GagPol acceptor: The 5' modified acceptor was tested along with the original acceptor by RNases to confirm their predicted structures (Figure 4-4). The two enzymes used were RNase A (purified from bovine pancreas) which specifically hydrolyzes RNA after C and U residues, and RNase T1 (purified from an *E.coli* strain carrying a cloned *Aspergillus oryzae* gene encoding this enzyme) which specifically hydrolyzes RNA after G residues. Figure 4-7 shows an autoradiogram of the assay conducted using T1 RNase and RNase A enzymes on the original and modified GagPol acceptor RNAs. With RNase T1 there was an absence of cleavage bands until position 33 from the 5' end of the original template; similarly, there were no cleavage bands from positions 149-158. This pattern is due to the protection conferred by the strong base pairing of the six Cs and one G between bases 149-158 at the 3' end with complementary bases at positions 19-28 at the 5' end. In contrast, an abundance of cleavage bands were seen at the same locations on the modified acceptor indicating the loss of the 5'-3' interaction. The assay repeated with RNase A enzyme revealed a similar pattern for both templates. The results were mapped to specific bases on both templates as shown in Figure 4-4. Notably, on both structures the predicted strong stem loop between bases 93-126 that carries the major pause site escaped digestion by both the enzymes and was left completely intact. Overall, these results illustrate that structures predicted by RNA folding programs are reliable and that the 5' modified GagPol acceptor closely matches with the GagPol donor RNA structure.

RNase protection assay of the 5' truncated and the 5' truncated mutant GagPol acceptors: In order to test the hypothesis that strand transfer is enhanced by low structure in the acceptor template, the 5' truncated (strongly structured) and 5' truncated mutated (weakly structured) GagPol acceptors were produced. The 5' truncated GagPol acceptor template was designed to retain the strong stem loop structure bearing the major pause site. This loop is between bases numbered 6-48 on Figure 4-5. The 5' truncated mutant GagPol acceptor was designed to eliminate the stem loop while retaining the pause site. RNase mapping was done with T1 enzyme to confirm the predicted structures as shown in Figure 4-8. For the 5' truncated acceptor only one cleavage band at position 38 was seen in the vicinity of the stem loop; all other bands were located outside of the stem loop. In contrast, the 5' truncated mutant was cleaved in a number of places indicating the weak nature of this template.
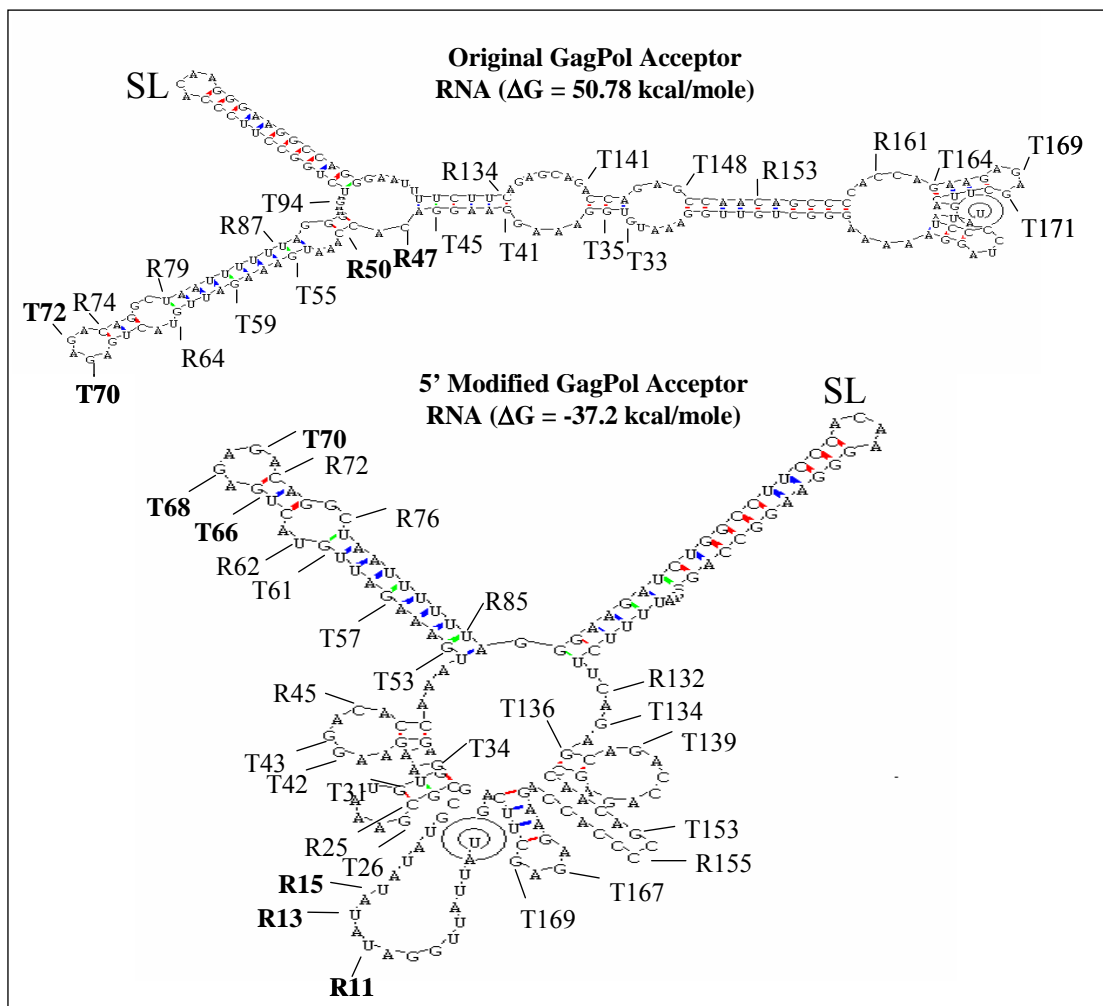
Figure 4-4: RNA structures of original and modified GagPol acceptors

Shown above are the predicted RNA structure of the original GagPol acceptor (bases 2009-2180 of the plasmid pNL4-3) and the 5' modified GagPol acceptor. The base encircled in two rings is the 5' end. The cleavage sites of T1 RNase enzyme are represented by the letter T along with the corresponding base numbers starting from the 5' end. The fonts 'T' and 'T' represent sites of moderate and intense cleavage, respectively. Similarly, the cleavage sites of RNase A enzyme are represented by the letter R along with the corresponding base numbers from the 5' end. SL marks the strong loop that carries the major pause site.
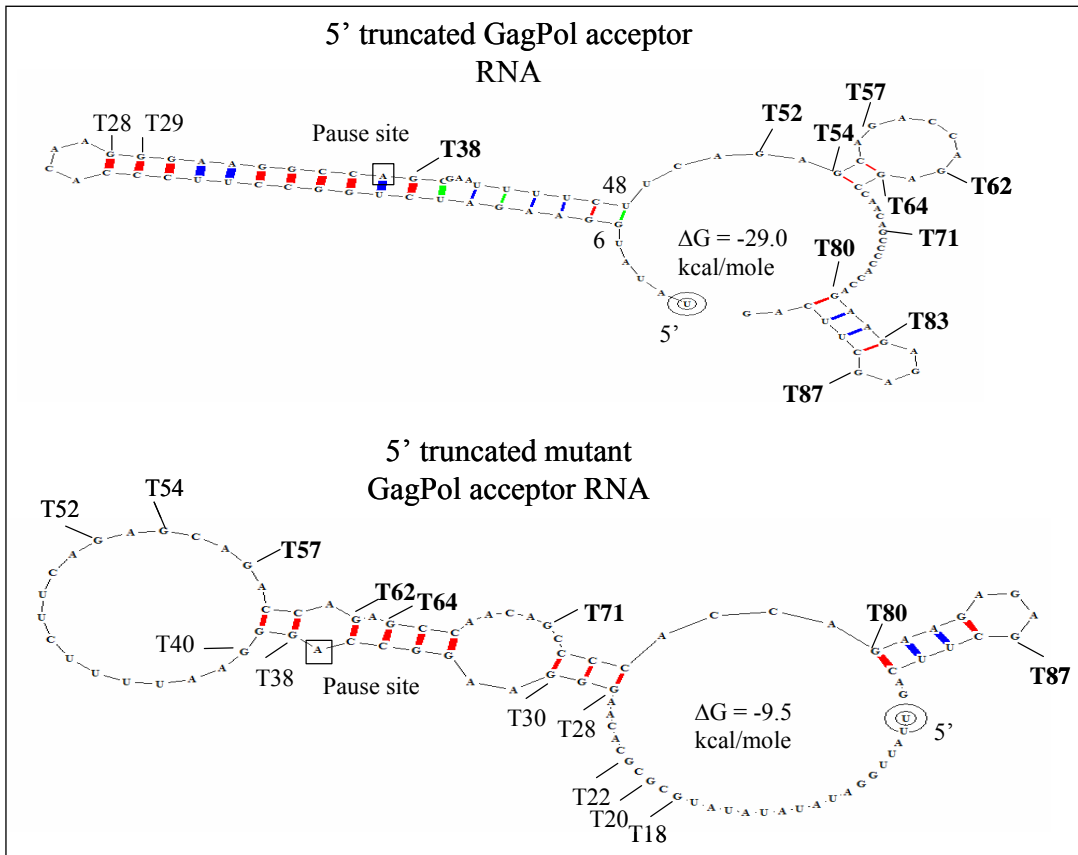
Figure 4-5: RNA structures of 5' truncated and 5' truncated mutant acceptors

Shown above are the predicted structures of the 5' truncated and 5' truncated mutant GagPol acceptor RNAs as indicated. The ΔG values of each structure are also shown. All other markings are the same as the legend of Figure 4-4.
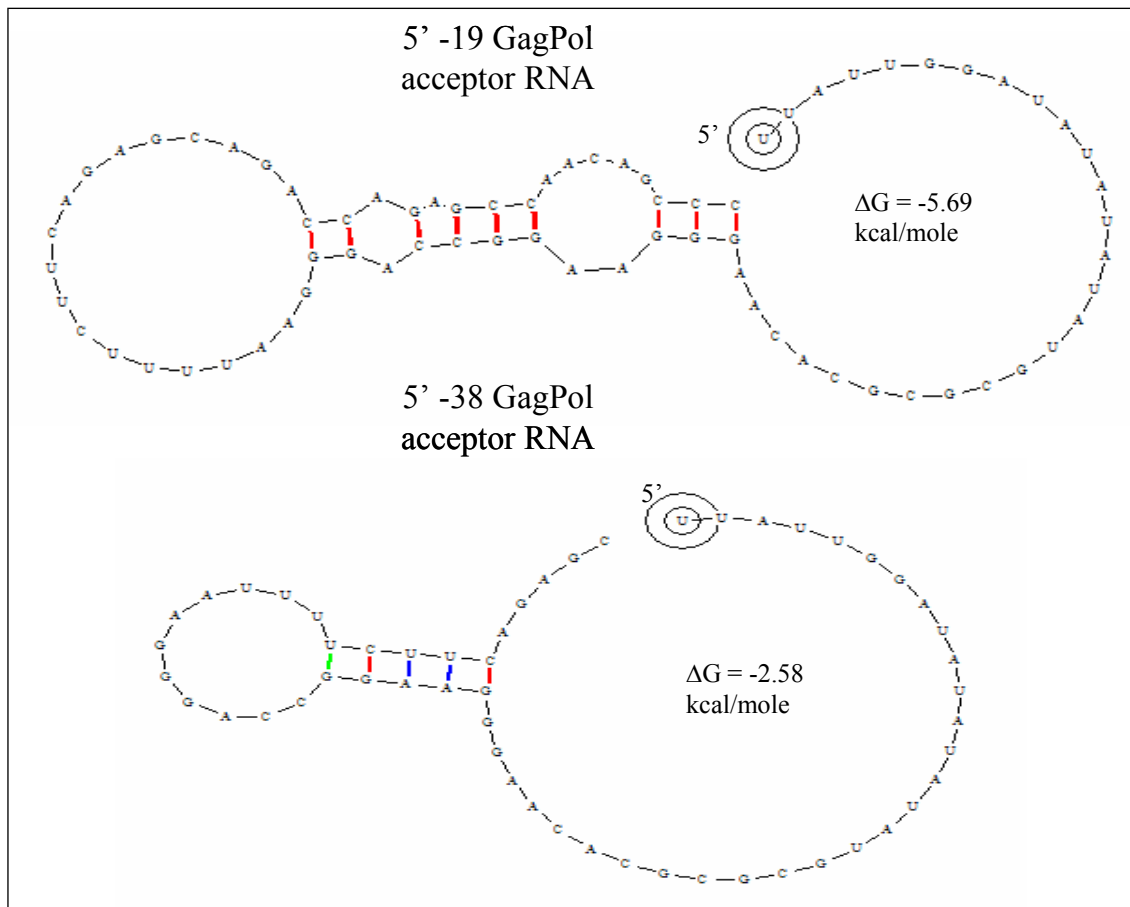
Figure 4-6: RNA structures of 3' -19 and 3' -38 GagPol acceptors

Shown above are the predicted structures of the 3' -19 and 3' -38 GagPol acceptor RNAs as indicated. The ΔG values of each structure are also shown. All other markings are the same as the legend of Figure 4-4.
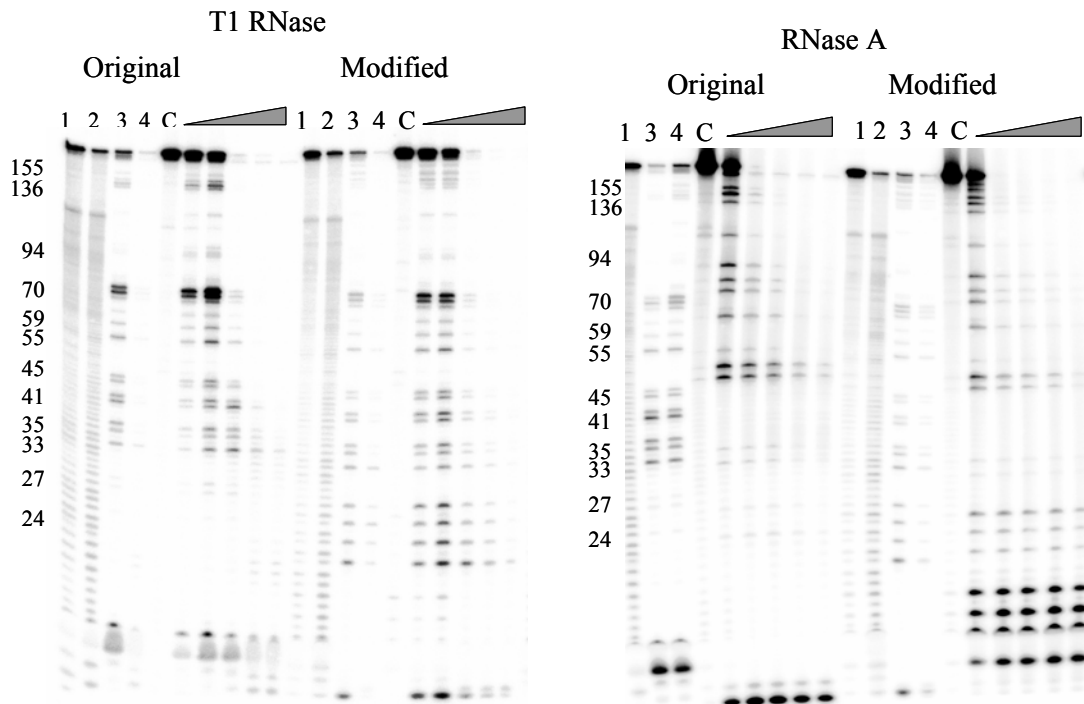
Figure 4-7: RNase mapping assays of original/modified GagPol acceptors

Shown is the RNase mapping assay conducted on the original and the 5' modified GagPol acceptor RNA sequences using T1 RNase and RNase A enzymes, as indicated. Lanes marked 1 and 2 are the base hydrolysis ladders conducted at 15 secs and 1 min. respectively, lanes marked 3 and 4 are the G ladders generated with 1 and 5 units of T1 RNase respectively, lanes marked 'C' are the control lanes containing radiolabeled RNA and no enzyme. The other lanes show the products of RNA digestion at increasing amounts of T1 RNase (from left to right, 0.5, 1, 5, 10 and 15 units, respectively) and of RNase A enzyme (1, 3, 5,7 and 9µunits, respectively) as indicated. The numbers on the left represent the base numbers of the G ladders.
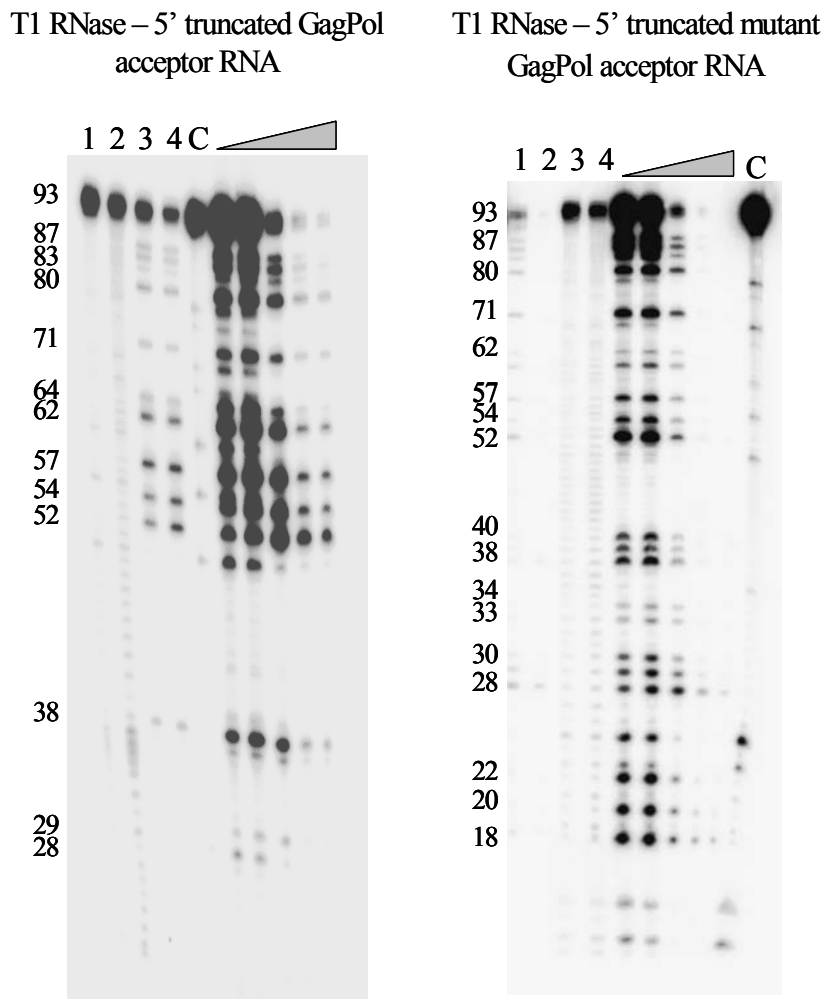
Figure 4-8: RNase mapping assays of 5' truncated and 5' truncated mutant GagPol acceptors

Shown is the T1 RNase enzyme mapping assay conducted on the 5' truncated and 5' truncated mutant GagPol acceptor RNA sequences using T1 RNase enzyme. All other markings are the same as the legend of Figure 4-7.

Effect of the five different GagPol acceptor RNAs on strand transfer: Strand transfer assays (time course experiments) were conducted using each of the five acceptors with the donor template. The 5' modified template did not show any significant differences in strand transfer when compared to the original GagPol acceptor. The pausing pattern and the transfer levels remained the same as the original GagPol acceptor (data not shown). Yet, this modified acceptor is actually a better test for 'authentic' transfer in this region than the original GagPol acceptor. This is because it has essentially the same structure as the donor. Within an infected cell the two copies of RNA genome would share the same structure. For this reason, this modified acceptor or truncations derived from it are used in the rest of the experiments.

Results for templates that carry truncations/mutations of the 5' end are shown in Figure 4-9. These acceptors were designed to eliminate transfers that occur downstream of the truncation in the full length acceptor. This allowed a careful analysis of events occurring upstream and including the major pause site (at position 77 on the GagPol donor). However, with these acceptors the transfer DNA products (113 bases) were shorter than the donor-directed products (175 bases) but longer than the paused DNA products (77 bases) (Figure 4-2). Since there were no other DNA products observed at position 113 in the absence of acceptor RNA, quantification of transfer was not difficult. It is important to note that despite the different folding pattern of the acceptors, they were identical in sequence for the first 70 nucleotides from the 3' end. Therefore transfers occurring before or even several bases after the pause site should not be affected by differences in homology between the donor and

acceptor. The enhancement of transfer with 5' truncated mutant was greater than with the 5' truncated RNA. The pause site faded rapidly within 16 minutes and was completely chased into transfer products by 64 minutes for the former, whereas, the paused DNA was still visible at 64 minutes with the latter. Although the transfer profile for the former in the absence of NC was higher than that of the latter in the presence of NC, there was significant stimulation of transfer by NC in both templates (Figure 4-10). These results were in agreement with the hypothesis that low acceptor structure favors transfer since the 5' truncated mutant lacks the strong stem loop found in the 5' truncated version. It is also a much weaker structure with a $\Delta G = -9.5$ kcal/mole as compared to $\Delta G = -29$ kcal/mole for the 5' truncated template. Overall, the results show that a destabilized acceptor RNA template is able to recombine better than a stable acceptor RNA. Also, structures downstream or 5' from the pause site probably play no role in transfers.

Numerous reports have proposed that pause driven transfers involve acceptor RNA and nascent DNA interactions upstream (or 3') from the pause site [119, 145]. In this case truncations at the 3' end of the acceptor template would be expected to have a negative effect on strand transfer. The 3' -19 and 3' -38 templates were constructed to test this supposition. Figure 4-3 shows lengths of overlap between 5' truncated mutant, 3' -19 and 3' -38 acceptors and the nascent DNA upstream from the pause site region. These acceptors are capable of forming hybrids corresponding to the lengths of overlap with nascent DNA. If transfers are occurring by acceptor invasion, then the shortest acceptor (3' -38) should be least efficient since it can form only a 19 base hybrid with the nascent DNA before the pause site. Results are shown

in Figure 4-11, Figure 4-12 and Figure 4-13. The truncated mutated GagPol acceptor was used as the basis for these experiments because of the high level of transfer observed with this acceptor. The strand transfer assays for 5' truncated mutant and the 3' truncation templates were compared to each other. Of all the acceptor templates examined so far, the destabilized 5' truncated mutant was the most effective in augmenting transfers. The 3' truncated versions were expected to diminish transfers but surprisingly the transfer profiles for all three templates were more or less equivalent (Figure 4-11, and Figure 4-12). There were small differences with the 3' -38 acceptor being the most effective enhancer of the three in the absence of NC and the intact truncated mutant being more effective with NC (Figure 4-13). These results indicate that sequences upstream of the pause site are also expendable in this mechanism of transfer. This argues that acceptor-facilitated type transfer, which postulated invasion of the DNA-donor complex upstream of the transfer point, does not occur at this pause site. Therefore dissociation of the DNA and donor after RNase H cleavage is the likely mechanism for transfer.
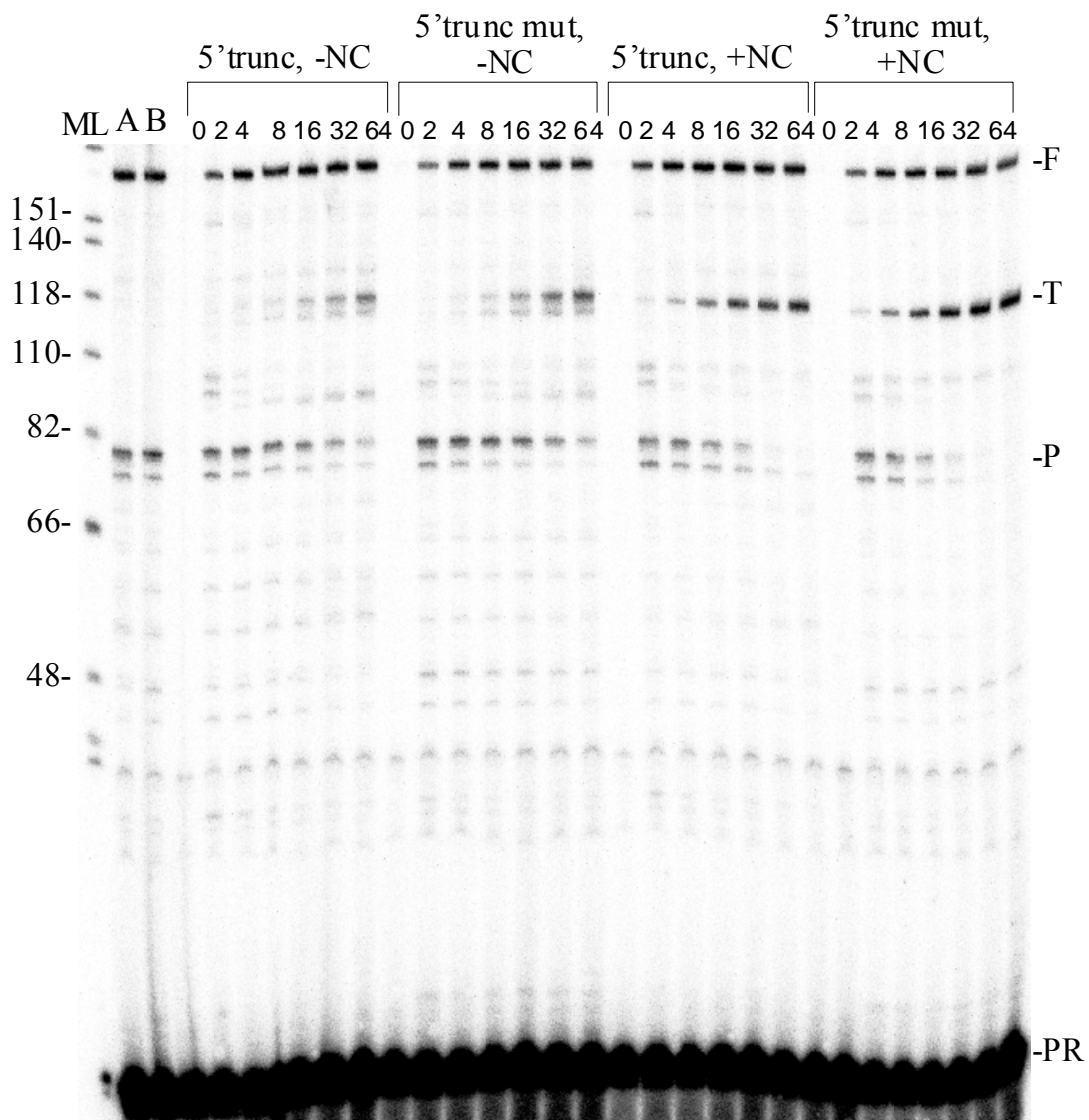
Figure 4-9: Time course assay of 5' truncated and 5' truncated mutant GagPol acceptors

Shown is an autoradiogram of time course assay performed using 5' truncated (5' trunc) and 5' truncated mutant (5' trunc mut) GagPol acceptor RNAs, as indicated. The leftmost lane 'ML' is the molecular ladder indicating the size (in nucleotides) of bands in the other lanes. The reactions were carried out either in the presence or absence of nucleocapsid protein (NC) as indicated above each set of lanes. GagPol Donor RNA was used in all reactions. Lanes marked A and B are reactions with no acceptor RNA and in the absence and presence of NC, respectively Each reaction was carried out at time points 0, 2, 4, 8, 16, 32 and 64 minutes as shown by the corresponding set of seven lanes from left to right. The transfer and full-length donor-directed DNA products, the pause site and the primers are indicated as T, F, P and PR respectively.

5' Truncated and 5' Truncated mutant GagPol Acceptor RNA templates

Legend:
- 5' Truncated GagPol Acceptor, -NC
- 5' Truncated GagPol Acceptor, +NC
- 5' Truncated Mutant GagPol Acceptor, -NC
- 5' Truncated Mutant GagPol Acceptor, +NC
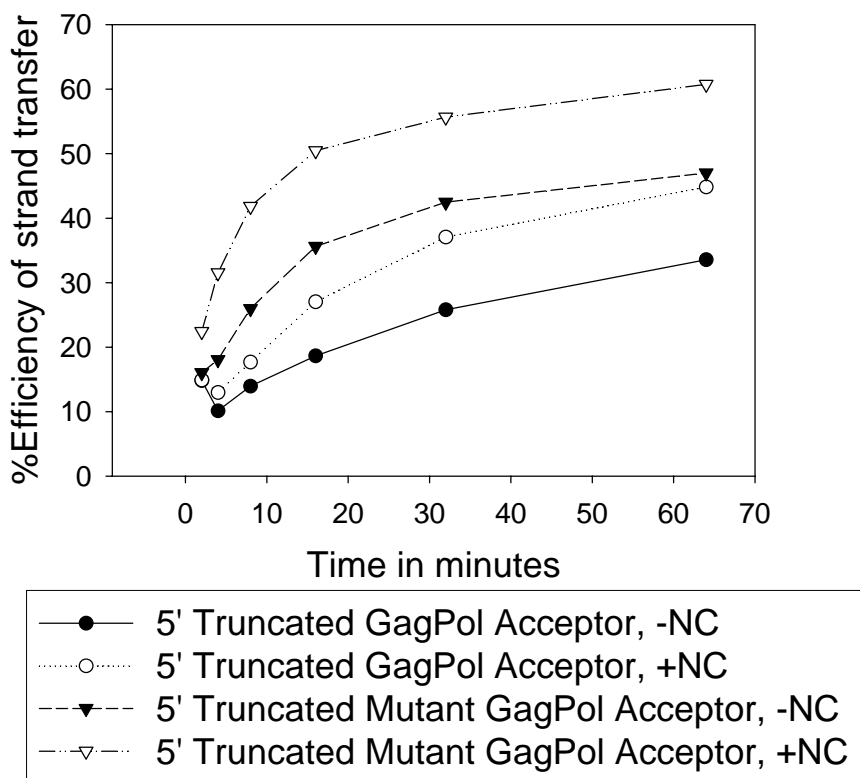
Figure 4-10: Efficiency of transfer in 5' truncated and 5' truncated mutant GagPol acceptors

Graph of efficiency of strand transfer *vs.* time for the 5' truncated and the 5' truncated mutant GagPol acceptor RNA, in the absence and the presence of nucleocapsid protein (-NC/+NC), as indicated. The percent transfer efficiency was defined as Transfer DNA products (T)/Transfer + Full-length donor-directed products (F), times 100 ((T/(T+F)) x 100).

-NC
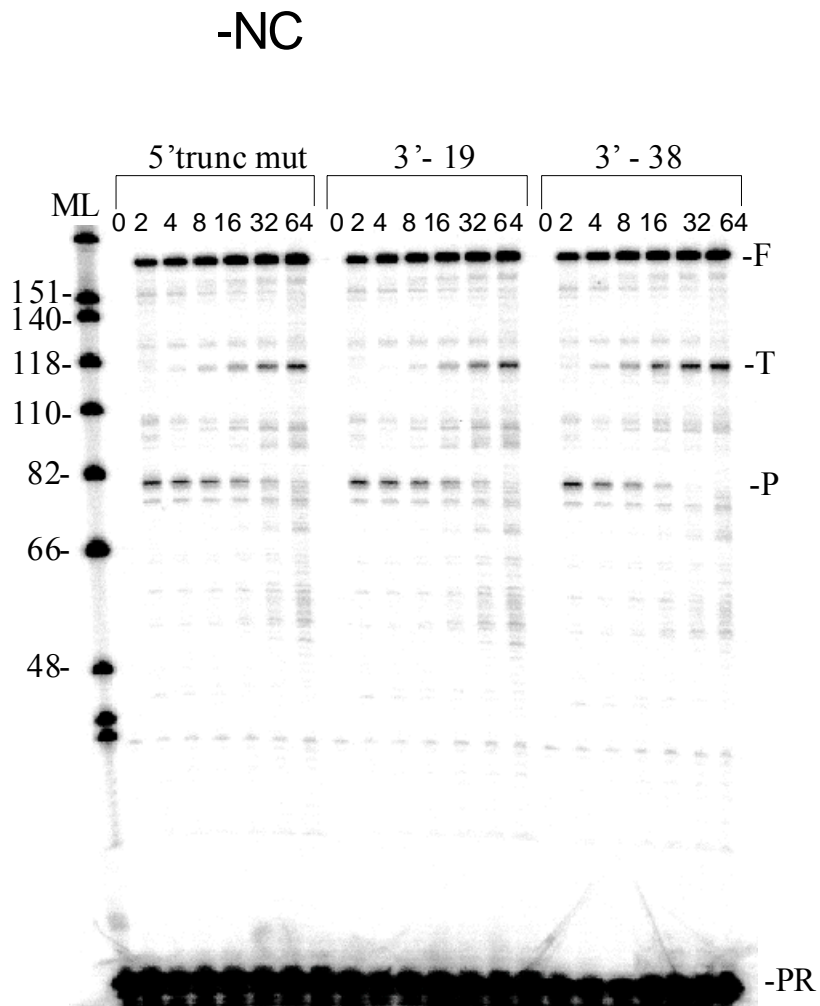
Figure 4-11: Time course assay of 5' truncated mutant, 3' -19 and 3' -38 GagPol acceptors, without NC

Shown above is the autoradiogram of a time course assay performed in the absence of nucleocapsid protein (-NC) with 5' truncated mutant (5'trunc mut), 3' -19 and 3' -38 GagPol acceptor RNAs, as indicated. All other markings are the same the legend of Figure 4-9
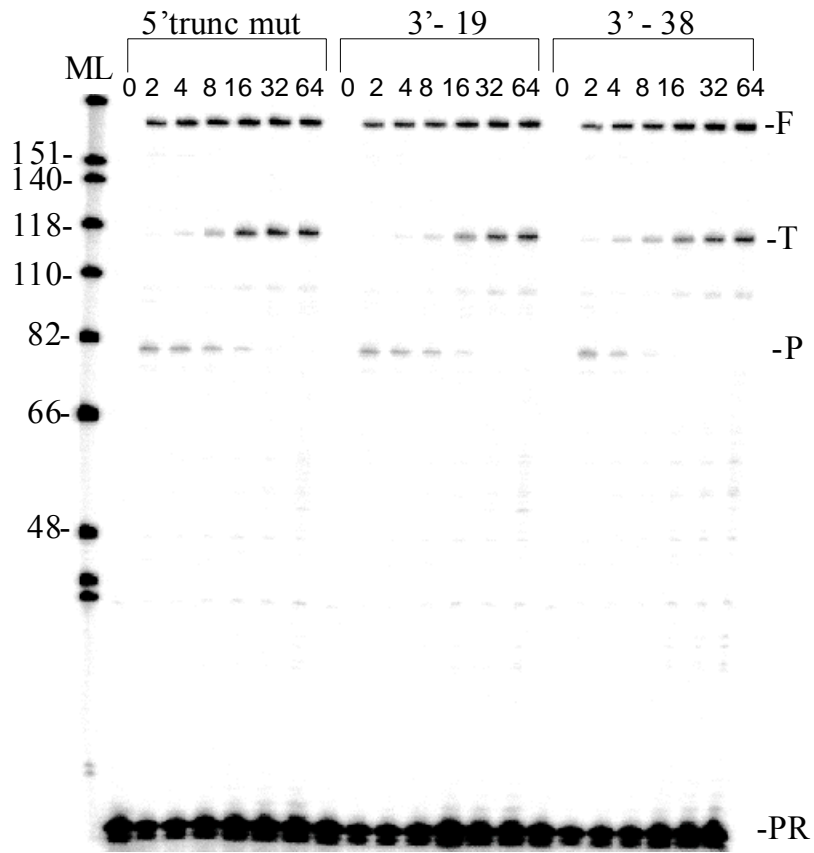
Figure 4-12: Time course assay of 5' truncated mutant, 3' -19 and 3' -38 GagPol acceptors, with NC

Shown above is the autoradiogram of a time course assay performed in the presence of nucleocapsid protein (+NC) with 5' truncated mutant (5'trunc mut), 3' -19 and 3' -38 GagPol acceptor RNAs, as indicated. All other markings are the same the legend of Figure 4-9

5' Truncated mutant, 3' -19 and 3' -38
GagPol Acceptor templates

Legend:
- ●——— 5' Truncated Mutant, -NC
- ○········· 3' -19, -NC
- ▼–––– 3' -38, -NC
- ▽–·–·– 5' Truncated Mutant, +NC
- ■— ■ — 3' -19, +NC
- □–·□·– 3' -38, +NC
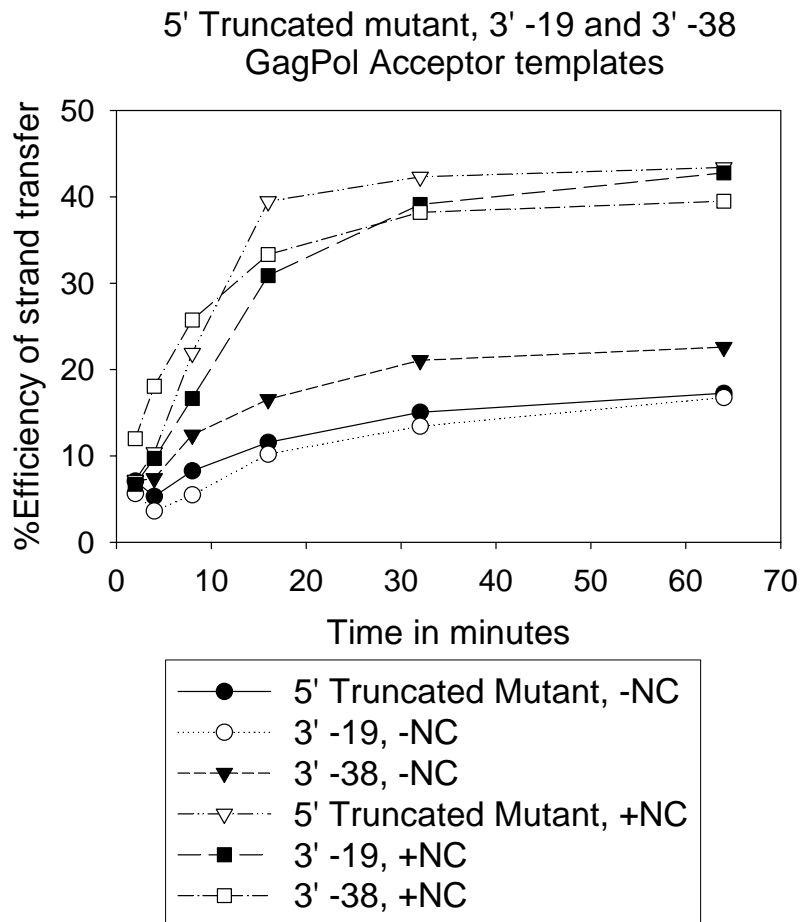
Figure 4-13: Efficiency of transfer in 5' truncated mutant, 3' -19 and 3' -38 GagPol acceptors

Graph of efficiency of strand transfer *vs.* time for the 5' truncated mutant, 3' -19 and 3' -38 GagPol acceptor RNA templates. The percent transfer efficiency was defined as Transfer DNA products (T)/Transfer + Full-length donor-directed products (F), times 100 ((T/(T+F)) x 100).

4.5 Discussion

The influence of acceptor RNA structure on strand transfer was analyzed by using 5' modified/truncated/mutated and 3' truncated versions of the GagPol acceptor with a non-variable GagPol donor in strand transfer assays. The 5'modified acceptor was structurally akin to the donor unlike the original acceptor used in Chapters 2 and 3. Even though it did not change the transfer profile it represents a more 'authentic' transfer, hence, all other versions were derived from it. Of all the 5' versions used the 5' truncated mutant was the weakest one (ΔG = -9.5 kcal/mole); it showed maximal efficiency of transfer. This outcome supports the notion that a destabilized acceptor recombines better than a stable one. In fact, a high level of transfer was seen for the destabilized 5' truncated mutant acceptor even without NC. This level was higher than that of the NC aided transfers of the stable 5' truncated acceptor (Figure 4-10). This difference in transfer levels with the two acceptors could also be a result of differential acceptor usage. In the case of the 5' truncated acceptor, transferred DNAs may not extend very well due to the presence of the strong pause site, resulting in non-productive transfer events. The usage of both acceptor RNAs would have to be compared to see if the low transfer level in 5' truncated acceptor is due in part to non-productive transfers.

The global implication of these results is that various HIV-1 subtypes can potentially recombine with each other rapidly in regions of their genomes that have differing thermodynamic stabilities. For example, HIV-1 subtypes A, E and F have *gag-pol* frameshift structures predicted to have lower thermodynamic stability than those of subtypes B, C and D [156]. This could potentially enhance intersubtype

recombination between pairs of viruses, especially if a highly structured and relatively low structure virus genome were co-packaged. If the former served as the donor and the latter the acceptor this should enhance recombination. The potential impact of these differences in stability on HIV-1 recombination remains to be determined.

During recombination in HIV, the nascent DNA has been proposed to transfer to the acceptor *via* two methods [54]. The acceptor invasion method requires active participation of the acceptor template. The DNA dissociation method occurs when nascent DNA falls off independently from the donor and later binds to the acceptor RNA. In the case of the former, it is vital for the 3' end of the acceptor to form a stable hybrid with the 5' end of the growing DNA strand upstream of the pause site. During pausing, NC would favor the more thermodynamically stable and longer acceptor-DNA hybrid over the donor-DNA hybrid which has been shortened by RNase H activity [65]. The 3' end terminus of the DNA then transfers to the acceptor either at or downstream from the pause site. In GagPol templates the pause site on the donor corresponds to position 57 from the 3' end of the acceptor RNA. The 3'-38 acceptor is the shortest of all the templates with only 19 homologous bases (57 less 38) at the 3' end that can bind to the nascent DNA before arriving at the pause site. It is improbable that an acceptor-DNA hybrid of 19 bases is adequate to efficiently displace the donor at a pause site by an acceptor facilitated mechanism (Figure 4-3). Yet, this acceptor was as effective as the longer 5' truncated mutant and the 3'-19 acceptors that can form 57 and 38 bases long hybrids, respectively, with the nascent DNA. This outcome advocates the possibility of the DNA dissociation method of

transfer which does not require long range acceptor-DNA interactions upstream from the pause site. The paused DNA is not extended in the absence of acceptor implying that it is no longer connected to the donor. The hybrid formed between the donor and paused DNA after RNase H cleavage is particularly weak. Initial cleavage by RT occurs about 18 bases upstream of the pause site on the donor. The 18 base hybrid thus produced has 11 A-U basepairs. Subsequent shortening of the hybrid by secondary RNase H cleavages would produce 14 and finally a 9 base hybrid. In the former 9 of the 14 basepairs are A-U and in the latter 6 of the 9 are A-U base pairs. A combination of the strength of the pause site and the weak hybrid could explain why the donor and DNA fall apart at this pause site and not others. These results are consistent with another *in vitro* report on DNA dissociation mode of transfer [54].

Chapter 5　Mapping of strand transfer in GagPol and Env RNA templates

5.1　Introduction

The assays conducted on various segments of the viral genome helped in identifying pause sites on highly structured RNA templates; the U3 3' LTR, GagPol and RRE regions(See Chapter 2). Strand transfer in these regions is likely to occur in a pause driven fashion. In contrast, weakly structured templates Env and PolVif showed a high rate of transfer despite the lack of major pause sites. The exact mechanism of transfer in these regions is unknown. The following experiments were designed with the goal of determining and comparing the points of strand transfer in one highly structured (GagPol) and in one weakly structured (Env) template. The GagPol template was chosen since its RNA structure has been well characterized by various investigators. The Env template is the weakest of all the templates that were analyzed (ΔG = -15 kcal/mol). The major point of transfer in GagPol is expected to be at or around the pause site that was earlier mapped to an A residue at position 77 from the 3'end of the donor template. In the case of the Env template it could be random or clustered around specific sequences or minor pause sites.

The experimental outline for mapping of transfer points is shown in Figure 5-1. Approximately equally spaced out point mutations were introduced into only the acceptor Env and GagPol RNA templates. In the case of GagPol templates, the modified GagPol acceptor was used according to reasons stated in Chapter 4. The mutations were carefully chosen to avoid causing drastic changes to the original structure of the acceptor templates. The mutated acceptors were then tested in strand

transfer assays to confirm that they exhibited transfer levels and pausing patterns that mimicked their wild type counterparts. The underlying premise of these experiments is that the nascent DNA that transfers from the wild type donor to the mutated acceptor and is subsequently extended, would carry some or all of the mutations depending on the point(s) of transfer. Consequently, sequencing of the transfer DNA from the GagPol and Env templates would expose the presence or absence of transfer point(s) between any two mutations.

An acceptor invasion model was proposed for transfers in the Env templates. In this case, acceptor template binds to nascent DNA then "zips" up the DNA toward the donor-DNA hybrid and displaces the donor (Figure 5-10). This model was tested by slowing the progression of the RT polymerase and thereby the rate of DNA synthesis. This would result in two events; (i) delayed polymerase activity would give more time to the acceptor RNA to displace the donor at more upstream locations within the regions of homology and (ii). the acceptor can 'steal' nascent DNAs that are still hybridized to the donor template. This should lead to a decrease in the level of full-length donor-directed products in reactions with acceptor. The possible occurrence of both these events was tested by conducting dNTP titration assays on both Env and GagPol templates. The results could aid in outlining a mechanism for strand transfer in weakly structured regions of the viral genome.
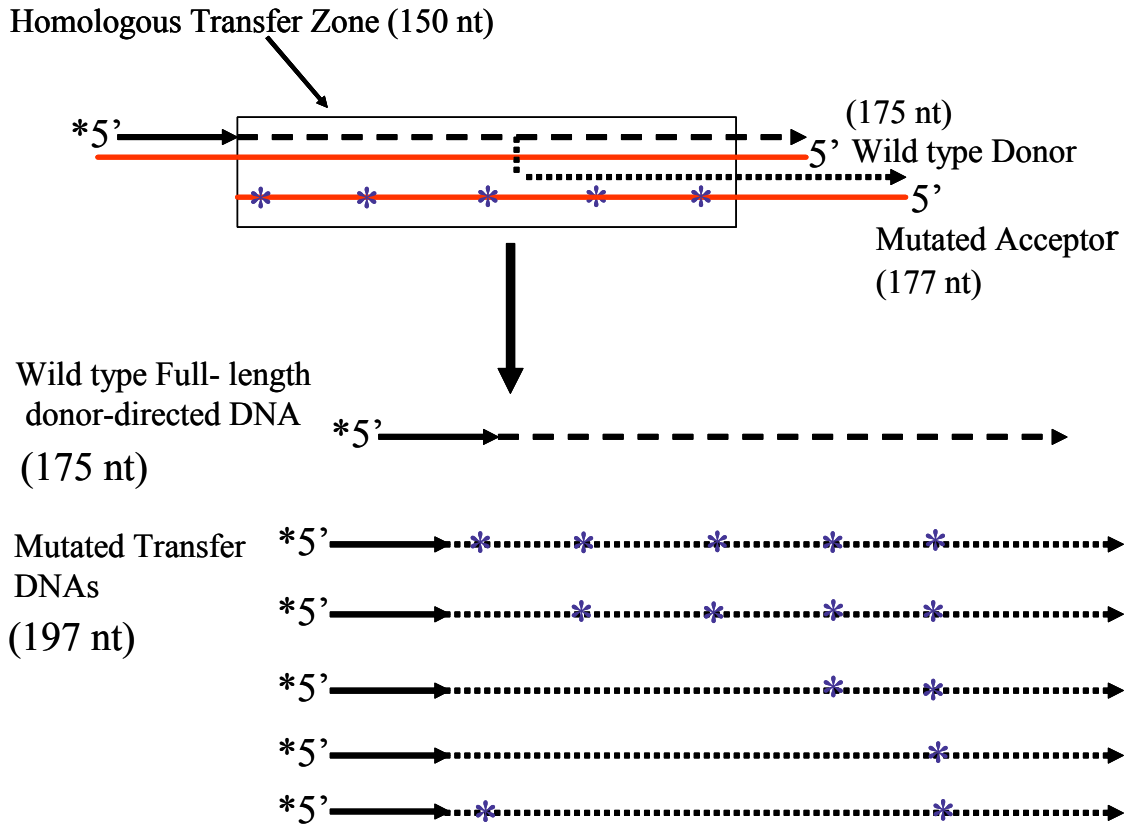
Figure 5-1: Outline of mapping assay

Shown above is the outline of the mapping assay used to locate the transfer points in Env and GagPol templates. The various substrates are indicated along with their lengths in nucleotides (nt). Each asterisk marks a mutated base on the acceptor RNA. The mutated transfer DNAs may carry one or all mutations depending on the point of transfer as illustrated under "Mutated Transfer DNAs". The last transfer DNA represents a double crossover event such that the first and the last mutations were picked up from the acceptor RNA.

5.2 Materials

All the mutant primers that were used to introduce mutations into the Env and GagPol acceptor templates are listed in Table 5-1. The primers were obtained from Integrated DNA Technologies, Inc. QuikChange site-directed mutagenesis kit was obtained from Stratagene. The PfuTurbo DNA polymerase used to introduce mutations in to pNL4-3 plasmid was from Stratagene. Plasmid pNL4-3, obtained from the NIH AIDS Research and Reference Reagent Program contains a complete copy of the HIV-1 provirus derived from strains NY5 and LAV [127]. Taq polymerase was from Eppendorf. SP6 RNA polymerase, DNase I-RNase free, and RNase-DNase free were from Roche Diagnostics. RNase inhibitor was from Promega. T4 Polynucleotide Kinase was obtained from New England Biolabs. Proteinase K was obtained from Kodak. Radiolabeled compounds: γ ^{32}P ATP was obtained from Amersham. Sephadex G-25 spin columns were from Amika Corp. Recombinant HIV-RT was provided to us by Genetics Institute (Cambridge, MA) (enzyme unit definition is as given in Section 2.2). Aliquots of HIV-RT were stored frozen at -70°C, and a fresh aliquot was used for each experiment. HIV-NC clone was a generous gift from Dr. Charles McHenry (University of Colorado). NC was purified and quantified as explained in Section 2.2. Aliquots of NC were stored frozen at -70°C, and a fresh aliquot was used for each experiment. Kits to prepare bacterial mini preps were obtained from Biorad. Topo TA cloning kits were from Invitrogen. All other chemicals were from Sigma or Fisher Scientific.

5.3 Methods

Mutation of the Env and GagPol acceptor templates: The mutant primers listed in Table 5-1 were used to introduce approximately equally spaced mutations into the previously used acceptor RNA templates from the *env* and *gag-pol* regions of the genome (Chapter 2). Five point mutations were introduced into the *env* region of the pNL4-3 plasmid. The mutations correspond to base numbers 32, 60, 91, 119, and 155 from the 3' end of the Env acceptor template (Figure 5-2). Similarly, four point mutations were introduced into the *gag-pol* region of the pNL4-3 plasmid that correspond to base numbers 45, 68, 89, and 105 from the 3'end of the GagPol acceptor template (Figure 5-3). The primers carrying each mutation were extended during temperature cycling by mutagenesis-grade PfuTurbo DNA polymerase (as per manufacturer's protocols). After temperature cycling, the product was treated with 1μl of Dpn I enzyme (provided in the kit). The Dpn I was used to digest the methylated parental DNA template and select for the newly synthesized mutated DNA. The plasmid DNA incorporating the desired mutations was then transformed into XL1-Blue cells. Mini preps were obtained using Biorad kits. Sequencing was done using primer 5'ctatctgttttaaagtggca 3' to check for mutation incorporation in the *env* region of the plasmid. Primer 5'ctgaagctctcttctggtgg 3' was used to check if the mutation was incorporated into the *gag-pol* region.

PCR amplification of DNA substrates: The PCR primers that were used to amplify the Env mutated acceptor template are;

5'-**gatttaggtgacactatag***tatat*aacacatctgtagaaat**a**a-3', and

5'-ctatctgttttaaagtggca-3'.

The base in bold and underlined is a mutated base at position 155 from the 3' end. The primers for the Env donor template are;

5' **gatttaggtgacactatag***atata*gtacaagacccaaca 3', and

5' ttgttctcttaatttgctag 3'.

The PCR primers that were used to amplify the GagPol mutated acceptor template are;

5' **gatttaggtgacactatag***tattattggatatatatatgcgcgc*gaaatgtggaaaggaagg 3', and

5' ctgaagctctcttctggtgg 3'.

The primers for GagPol donor template are;

5'**gatttaggtgacactatag***tcaaa*gaaatgtggaaagga 3', and

5' ttgttgtctctaccccagac 3'.

The sequences in bold are the SP6 promoter sequences that allow transcription of the DNA by SP6 RNA polymerase. The italicized sequences on the donor are five non-HIV-derived nucleotides that disrupt the homology between the donor and acceptor at the 5'end. This prevents strand transfers from the end of donor and restricts transfers to internal regions. In the case of the primers for the GagPol mutated acceptor template, there are 25 italicized nucleotides. These were designed specifically to make the GagPol acceptor and donor templates look structurally alike.

113

The PCR cycling parameters and the protocols for DNA/RNA preparation and purification are explained in Section 2.3.

RNA-DNA hybridization: DNA primers that bound specifically to the wild type Env and GagPol donor RNA transcripts were ^{32}P-labeled at the 5' end with T4 polynucleotide kinase according to the manufacturer's protocol. Each of the donor RNAs was hybridized to a complementary labeled primer by mixing primer:transcript at approximately 5:1 ratio in 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 80 mM KCl, and 0.1 mM EDTA (pH 8.0). The mixture was heated to 65°C for 5 min and then slowly cooled to room temperature. Time course reactions and gel electrophoreses were performed as explained in Section 2.3. The dNTP titration experiments were carried out on Env and GagPol templates, by conducting time course reactions in the presence of 1 and 100 μm dNTPs

PCR amplification of Transfer DNA products: In the time course reactions the Env transfer DNA products (197 nucleotides) obtained at 32 minutes in the absence and in the presence of NC were gel purified on denaturing 8% polyacrylamide gels and located by autoradiography. The DNAs were excised and eluted overnight in a TE buffer (10 mM Tris-HCL (pH 8.0), 1 mM EDTA (pH 8.0)). The eluate was separated from the gel by centrifugation and subsequent filteration through a 0.45μ disposable syringe filter. The DNAs were recovered by precipitation in ethanol with 300 mM sodium acetate. The recovered DNA was further PCR amplified with primers; 5'**gatttaggtgacactatag***tatat*aacacatctgtagaaat<u>a</u>a 3' and 5' ttgttctcttaatttgctag 3'. The resultant DNA was ligated into Topo vector which was used to transform Top10' *E.coli* competent cells (as per manufacturer's protocol). Only the white and

114

pale blue colonies were picked. Minipreps were prepared using a Biorad miniprep kit. DNA was sequenced using M13 reverse primer. The results are in shown in Figure 5-4 and explained in the Results section. Similarly, the transfer DNA products (195 nucleotides) were purified for the GagPol templates and PCR amplified with the primers; 5'**gatttaggtgacactatag**_tattattggatatatatatgcgcgc_gaaatgtggaaaggaagg 3' and 5' ttgttgtctctaccccagac 3'. The recovered DNA was used to transform Top10' *E.coli* cells and processed as described above. The sequencing results of the positive colonies are shown in Figure 5-5 and explained in the Results section. The Env transfer DNA products from dNTP titration experiments were also PCR amplified and sequenced as mentioned above. The sequencing results are shown in Figure 5-4.

Table 5-1: Mutant primers (Env and GagPol regions)

Region[a]	base[b]	sequences of primers in the 5' and 3' orientations[c]
Env acceptor, 7079-7250	32	5'gcacattgtaacattagtag*t*gcaaaatggaatgccac3' 5'gtggcattccattttgc*a*ctactaatgttacaatgtgc3'
	60	5'ggaaaaataggaaata*a*gagacaagcacattg3' 5'caatgtgcttgtctc*t*tatttcctattttttcc3'
	91	5'gaggggaccagggaga*t*catttgttacaatagg3' 5'cctattgtaacaaatg*a*tctccctggtcccctc3'
	119	5'caacaatacaagaaaaag*g*atccgtatccagagggg3' 5'cccctctggatacggat*c*cttttttcttgtattgttg3'
	155	5'ctgaacacatctgtagaaat*a*aattgtacaagacccaacaac3' 5'gttgttgggtcttgtacaatt*t*atttctacagatgtgttcag3'
GagPol acceptor, 2031-2180	45	5'cagggaattttct*a*cagagcagaccag 3' 5' ctggtctgctctg*t*agaaaattccctg 3'
	68	5'aagatctggccttccccac*t*agggaaggccagggaatt3' 5'aattccctggccttccct*a*gtgggaaggccagatctt 3'
	89	5'caggctaattttttta*c*ggaagatctggccttc3' 5'gaaggccagatcttcc*g*taaaaaattagcctg-3'
	105	5'gaaagattgtactgagag*t*caggctaattttttagg-3' 5'cctaaaaaattagcctg*a*ctctcagtacaatctttc3'

a- The numbers below each acceptor refer to the sequence numbering of the provirus in the pNL4-3 plasmid.

b- These numbers represent the position of each point mutation from the 3' end of the Env or GagPol acceptor RNA as indicated.

c- These are primer pairs that are complementary to opposite strands of pNL4-3 plasmid at base numbers (shown above) located within the Env or the GagPol RNAs. The underlined letters are the mutated bases.
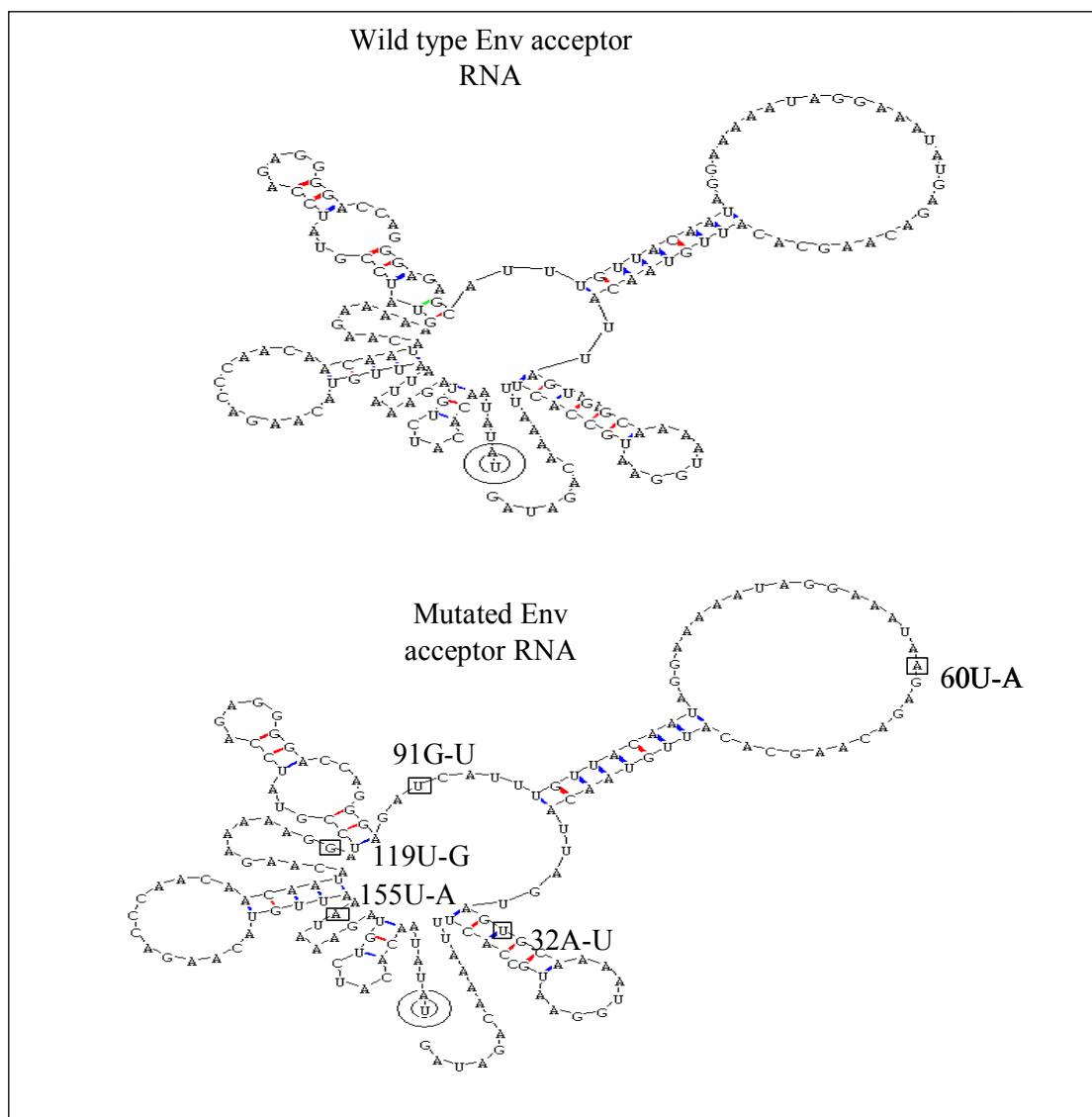
Figure 5-2: Wildtype and mutated Env acceptor RNAs.

RNA folding programs were used to predict the weakly structured wild type and mutated Env acceptor RNA templates (177 bases each). The base enclosed in two rings is the 5' end. '32A-U' represents a wild type A residue at location 45 from the 3' end of the acceptor that was mutated to a U residue (enclosed in the box). Four other mutated bases are shown in a similar way.
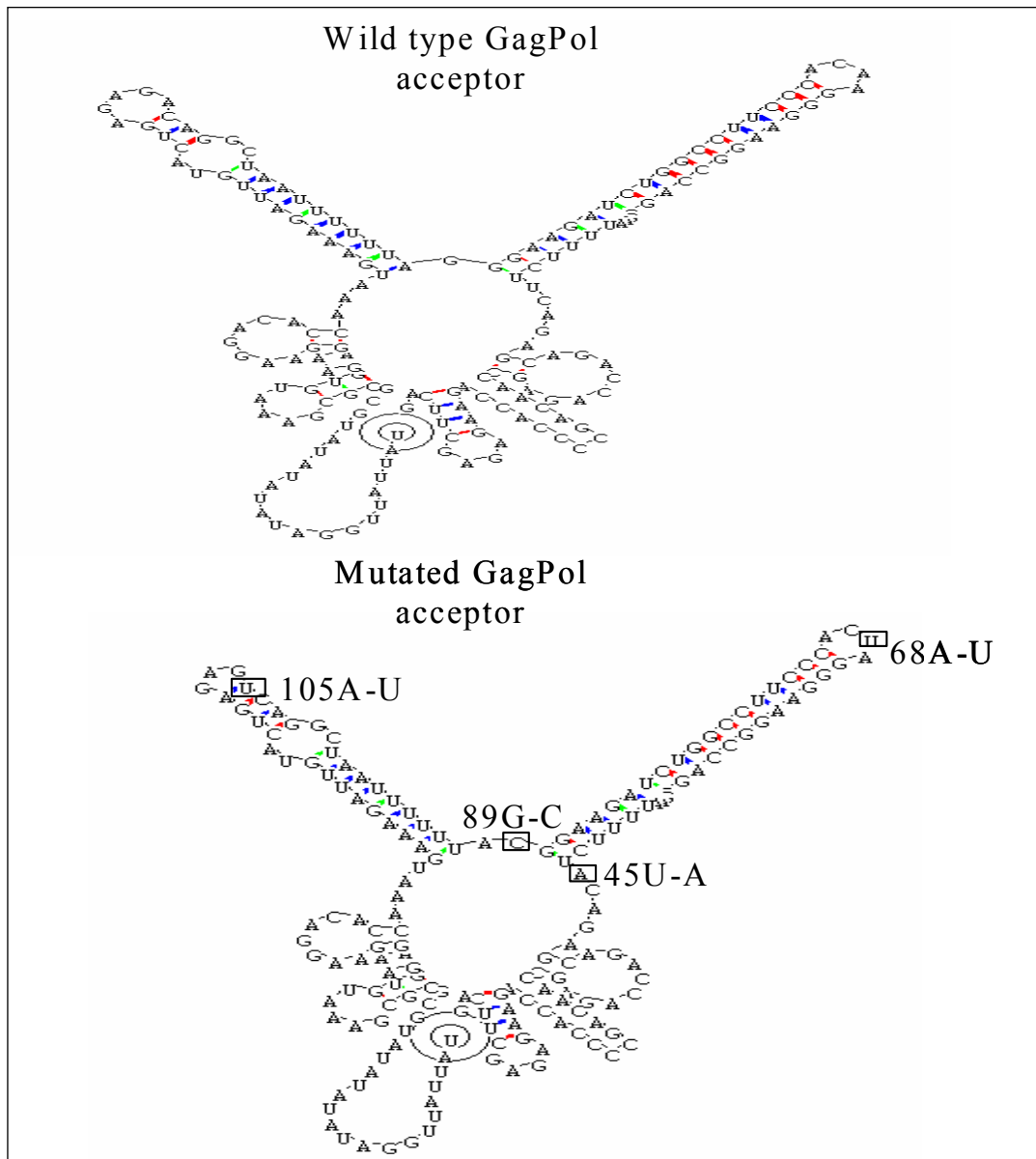
Figure 5-3: Wild type and mutated GagPol acceptor RNAs.

RNA folding programs were used to predict the structure of the wild type and mutated GagPol acceptor RNA templates (177 bases each). The base enclosed in two rings is the 5' end. '45U-A' represents a wild type U residue at location 45 from the 3' end of the acceptor that was mutated to an A residue (enclosed in the box). Three other mutated bases are shown in a similar way.

5.4 Results

Construction of Env and GagPol substrates: The Env donor RNA template is non-mutated and 175 nucleotides in length. The Env acceptor RNA template is 177 nucleotides in length with five equally spaced point mutations that correspond to base numbers 32, 60, 91, 119, and 155 from the 3' end; the last mutation at base 155 lies just outside the region of homology. The mutations were carefully selected so as to retain the wild-type acceptor structure (Figure 5-2). The GagPol donor RNA template is non-mutated and 175 nucleotides in length. The GagPol acceptor template is also 175 nucleotides in length with four equally spaced point mutations that correspond to base numbers 45, 68, 89, and 105 from the 3'end of the template (Figure 5-3). All four mutations lay within strong stem loops of the region of homology. The mutations were all made in the 175 nucleotide GagPol modified acceptor instead of the 177 nucleotide original acceptor because of the reasons previously stated in Chapter 4.

Time course experiment results: The Env and the GagPol donor-mutated acceptor pairs were analyzed by time course reactions as described in Chapter 2. The results matched with those of the wild type acceptors indicating that the introduced mutations did not have any drastic effects on the overall system (data not shown).

Mapping of the point of transfer on Env/GagPol substrates: DNA synthesis was initiated from a 5' end-labeled DNA primer that was specifically designed to bind only to the 3' end of the Env/GagPol donor RNAs (Figure 5-4, and Figure 5-5). Strand transfer can occur over the transfer zone, which is the region of homology

between the donor and acceptor RNAs. Primer extension to the end of the donor produced a 175-base full-length donor-directed DNA product that lacked mutations. Strand transfers resulted in 197-base transfer DNA products that carried a combination of mutations depending on the location of transfer. For example, on the Env templates if a single transfer event occurred before the first mutation at base 32 and synthesis continued till the end of the acceptor then transfer DNA would incorporate all five mutations that are on the acceptor. If it occurred between mutations 32 and 60 then the transfer DNA would have only four of the mutations that are downstream from the base 32 mutation. These two scenarios involve single crossovers. The results vary for double or multiple crossovers wherein the transfer DNA jumps back and forth between the donor and acceptor templates. This system could map the location of transfer between any two mutations. Transfer DNA products accumulate to significant amounts at 32 minutes in a time course reaction, hence products from this time point in the presence and absence of NC were carefully purified and amplified by PCR. Transfer DNAs were devoid of any contamination by full-length DNAs because of specially designed PCR primers. The amplified transfer DNAs were cloned and sequenced. The sequencing results for the Env templates are shown in Figure 5-4. In the presence of NC, nearly 46% of transfers occurred between mutations at bases 119 and 155. Another 34% of transfers were at bases between 91 and 119. In the absence of NC, many transfer points were found at the earlier mutations at positions 32 and 60, still the majority of transfers were concentrated towards the last mutations at 91 and 119. The results indicate that transfer points in weakly structured regions are located towards the end of the region

of homology. Transfer DNAs at 8 minutes with NC, 32 minutes with and without NC were sequenced for the GagPol templates. The early time point was chosen to test for any differences in transfer patterns between early and late time points. Most transfers (78% for 8 minutes +NC and 94% for 32 minutes +NC) occurred between mutations at bases 45 and 68, in other words, only mutations 68, 89 and 105 were present in a majority of the sequenced clones. The data shows that the preferred transfer site in GagPol lies at or just after the pause site at base 57. These results are in contrast to those of the Env templates where transfers toward the end of the region of homology are predominant. Clearly, the GagPol template depends on its pause site for a majority of transfer events. Also, the results from Chapter 4 have established a donor dissociation mode of transfer in this case. On the basis of the mapping results for the Env template an acceptor invasion mode of transfer was proposed, which was tested by dNTP titration assays (explained below).

dNTP titration assays of Env and GagPol templates: In these assays, low concentrations of dNTPs were used to slow the progression of RT polymerase. Time course reactions with 1 and 100 μM dNTPs were conducted on Env and GagPol templates. Nucleocapsid protein was used in all the reactions. The GagPol donor has been shown to dissociate independently of the acceptor (in Chapter 4). Results from the GagPol templates would serve as a negative control in this assay. Mapping of the Env transfer DNA products at 1 μM dNTPs and 32 minute time point was done as described earlier (Figure 5-4). A shift of transfer points from between mutations 119 and 155 at the end of the homologous regions to mutations at upstream locations was observed. This is in agreement with the prediction of event (i) as explained in the

Introduction. Mapping of GagPol transfer DNA at 1 µM dNTPs was not done, since there was no effective "chasing" of the pause site or production of optimum levels of transfer DNA.

Figure 5-6, and Figure 5-7 show the autoradiogram of dNTP titration assays conducted on Env and GagPol templates, respectively. The efficiency of transfer for Env templates rose from about 50% at 100 µM concentration to over 70% at 1 µM concentration and 64 minutes time point (Figure 5-8). In reactions with acceptor there was a pronounced increase in transfer DNA at the expense of donor-directed DNA products, over time. In contrast, the overall levels of both donor-directed and transfer DNA were very low at 1 µM dNTPs in the GagPol templates. The efficiency of transfer for GagPol templates was about 50% at 64 minutes and at 1µM concentration (Figure 5-8). This was only a slight increase over the 40% level at 100 µM concentration at the same time point. In this template, giving more time to the acceptor to bind to the nascent DNA did not enhance transfer, once again suggesting a DNA dissociation mode of transfer (Figure 5-9).

The levels of donor-directed DNA for Env and GagPol was at 1 µM dNTPs and either in the presence or absence of acceptor RNA was compared in Figure 5-8. There was a large drop in the donor-directed DNA levels for Env templates in the presence of acceptor RNA. A similar drop was not present in the GagPol templates, these results are consistent with the prediction of event (ii) as explained in the introduction.
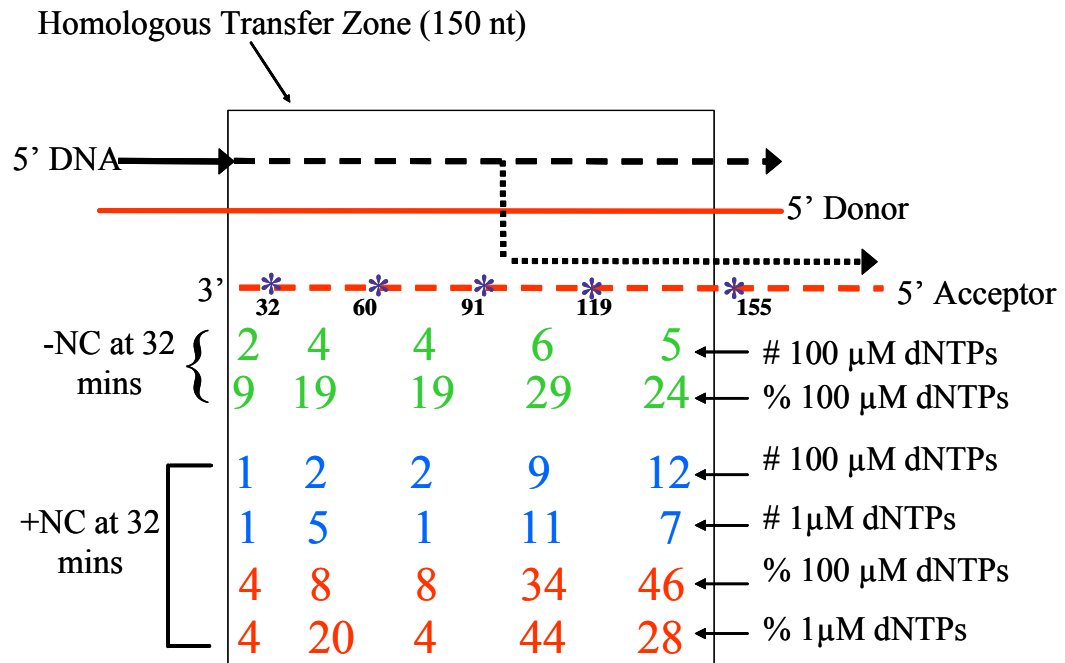
Figure 5-4: Sequencing results for Env template.

Shown above are schematics of the wild type donor and mutated acceptor Env RNA templates. The mutations (marked by asterisks) are located at positions 32, 60, 91, 119 and 155 from the 3' end of the acceptor. The number (#) and percentage (%) of transfer DNAs that transferred between any two mutations at 32 minutes in the presence (+NC) or absence (-NC) of nucleocapsid protein, either with 100 or 1 μM of dNTPs are shown within the box.
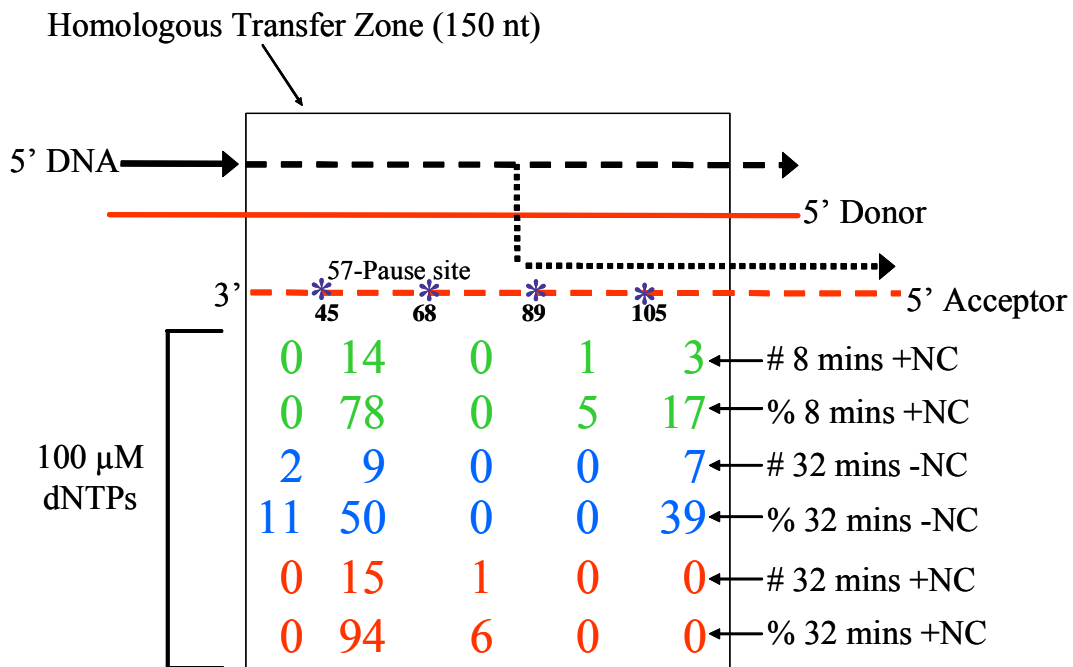
Figure 5-5: Sequencing results for GagPol template.

Shown above are schematics of the wild type donor and mutated GagPol acceptor RNA templates. The mutations (marked by asterisks) are located at positions 45, 68, 85 and 109 from the 3' end of the acceptor. The pause site at position 57 on acceptor is also shown. The number (#) and percentage (%) of transfer DNAs that transferred between any two mutations at 8 or 32 minutes, either in the presence (+NC) or absence (-NC) of nucleocapsid protein are shown within the box. Transfer DNAs were obtained from time course reactions conducted with 100 μM of dNTPs.
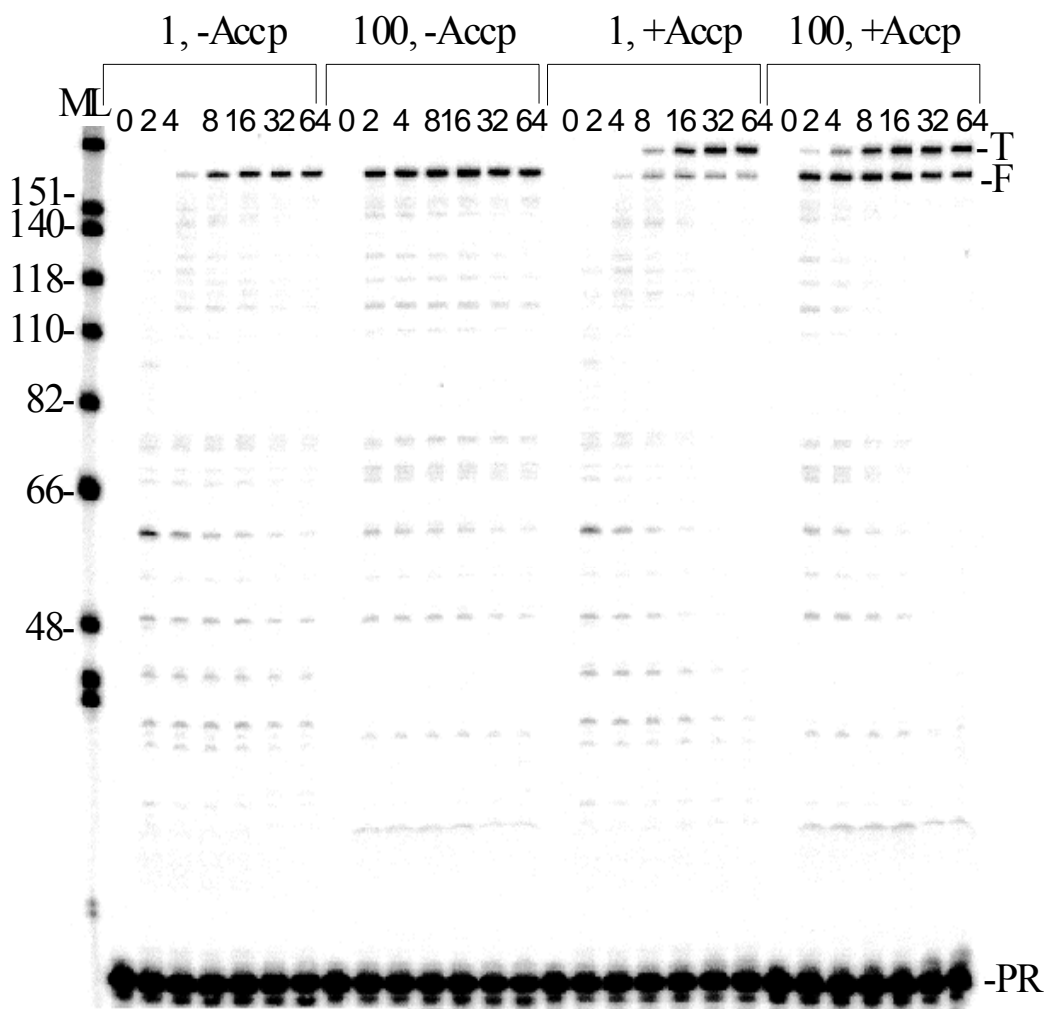
Figure 5-6: dNTP titration of Env templates

Shown above is an autoradiogram of dNTP titration assay performed on Env templates. ML is the molecular ladder indicating the positions of the bands in the other lanes. The reactions were carried out in the presence or absence of Env acceptor RNA (-/+Accp) and with 1 or 100 μM of dNTPs as shown above each set of lanes. Nucleocapsid protein was used in all the reactions. Each reaction was carried out at time point 0, 2, 4, 8, 16, 32 and 64 minutes as shown by the corresponding set of seven lanes from left to right. The transfer, full-length donor-directed DNA products and primers are indicated as T, F and PR, respectively.
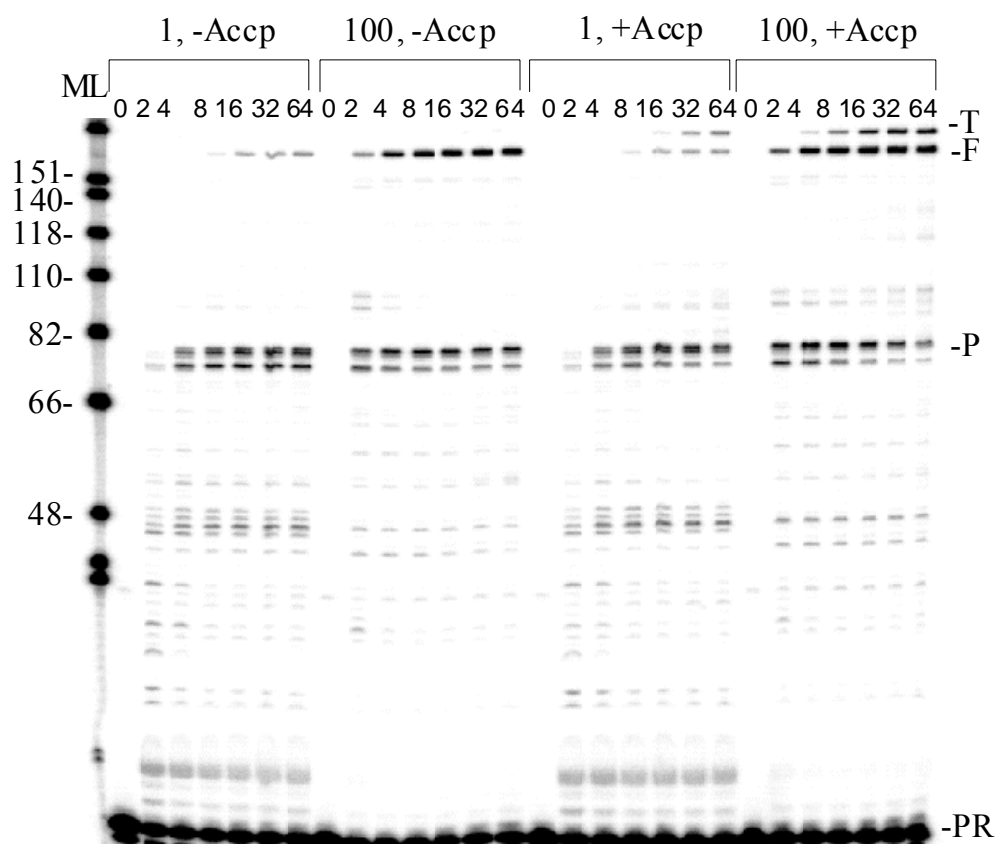
Figure 5-7: dNTP titration of GagPol templates

Shown above is an autoradiogram of dNTP titration assay performed on GagPol templates. All markings are the same as legend of Figure 5-6.
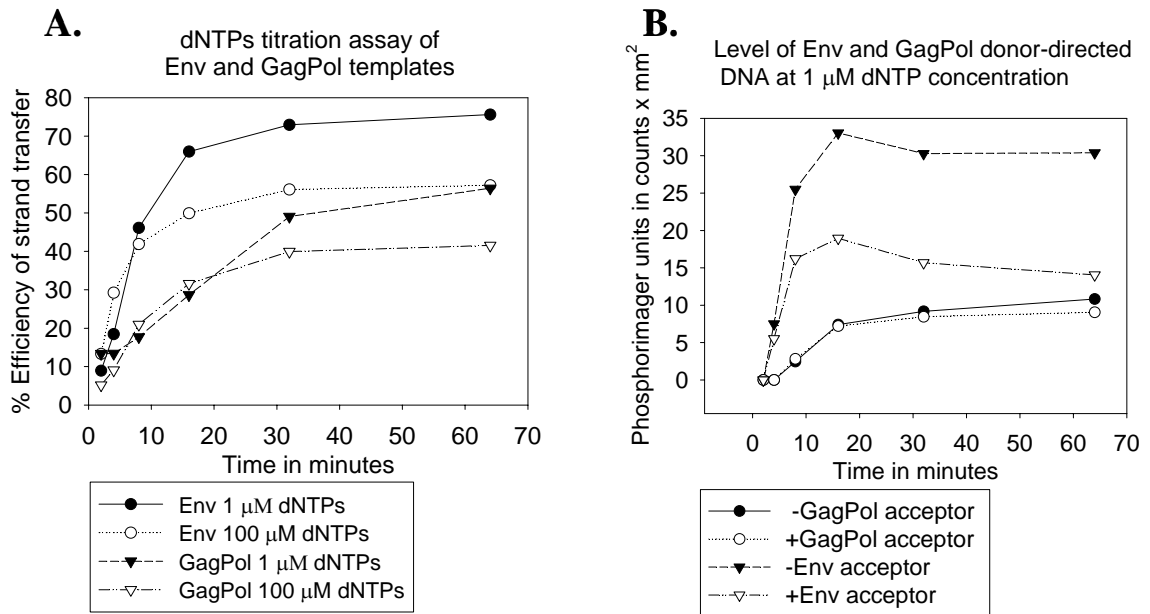
Figure 5-8: Graphs of dNTP titration assay and levels of donor-directed DNA.

Graphs were derived from dNTP titration experiments conducted on Env and GagPol templates with 1 or 100 µM dNTP concentrations as indicated. Graph A shows the percent transfer efficiency that is defined as Transfer DNA products (T)/Transfer + Full-length donor-directed products (F), times 100 ()T/T+F x 100). Graph B shows the levels of donor-directed DNA in the presence of 1 µM dNTP concentration, obtained by phosphorimager analyses
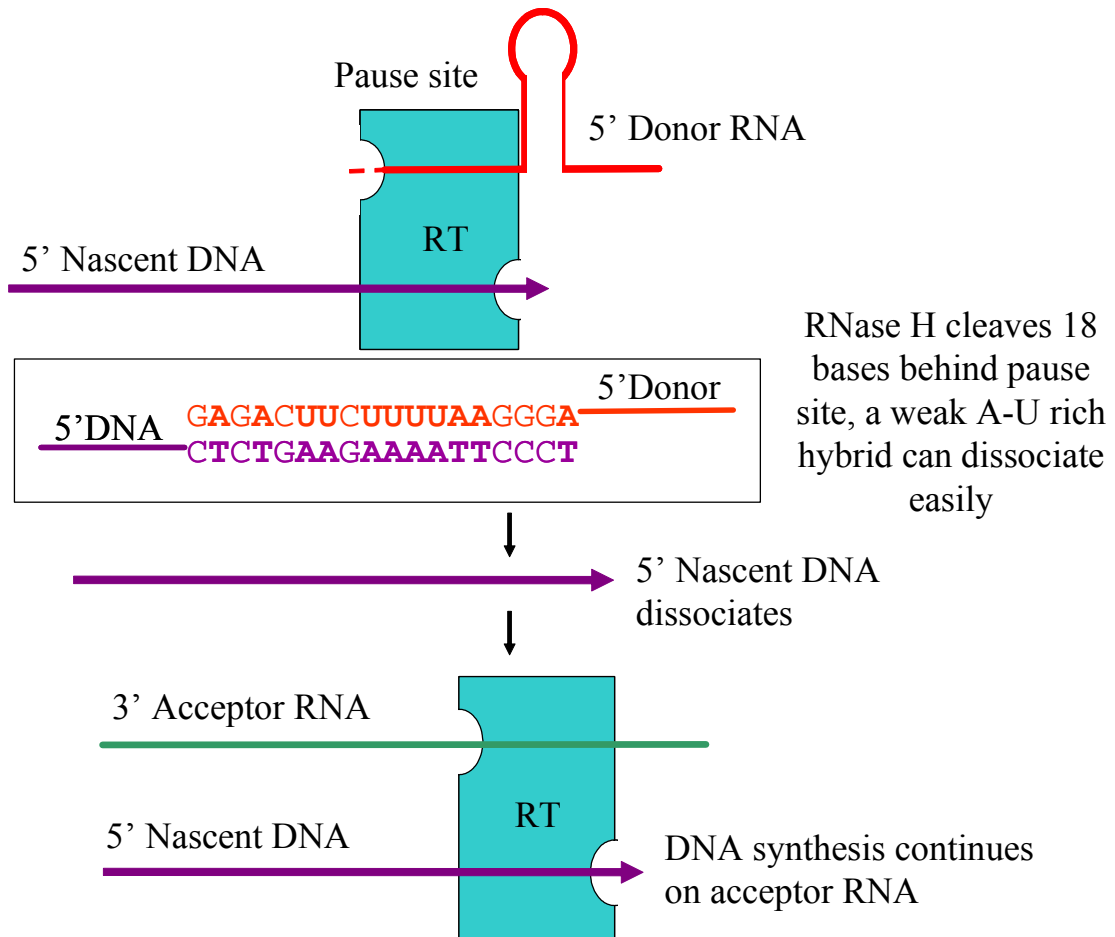
Figure 5-9: DNA dissociation model for GagPol templates

Shown above is the pause-induced DNA dissociation model for transfers in GagPol templates. The red lines and letters show the donor RNA. Broked red lines are cleavage products of RNase H activity. The dark purple lines and letters are the nascent DNA. At the pause site at the stem loop, nascent DNA dissociates and then binds to acceptor RNA (green line). RT is the reverse transcriptase enzyme.
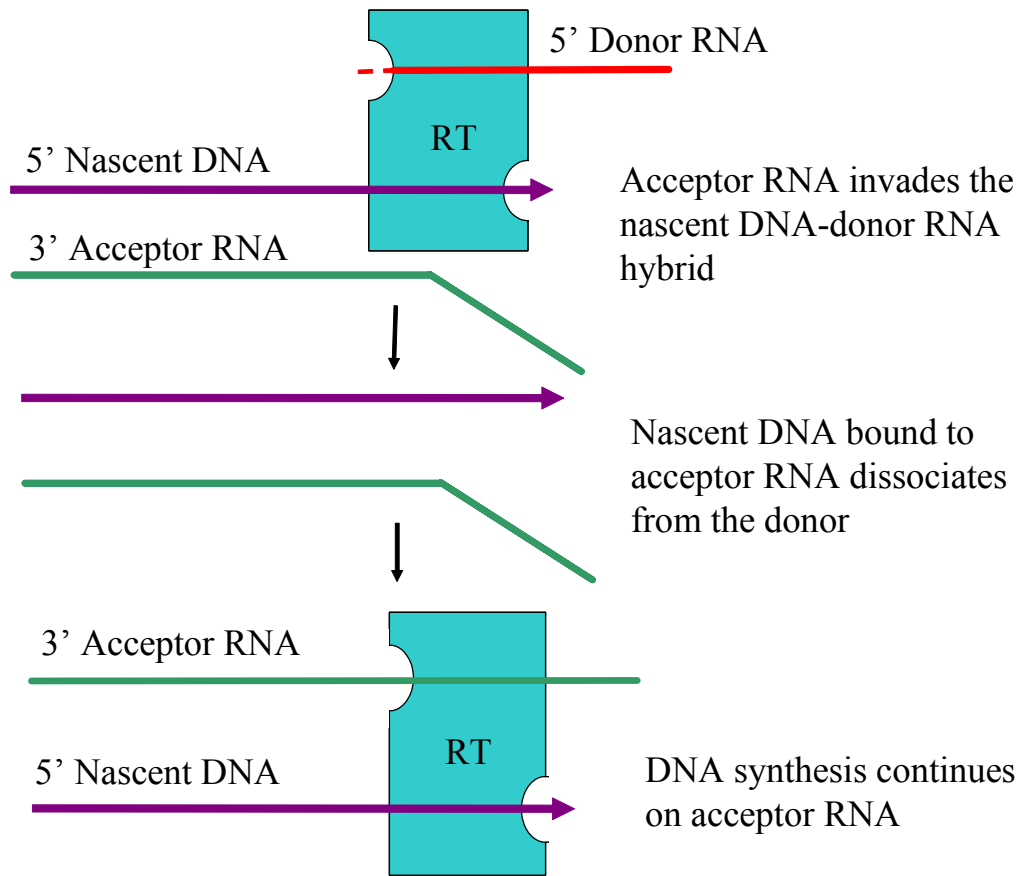
Figure 5-10: Acceptor invasion model for Env template

Shown above is the pause-independent, acceptor invasion model for transfers in Env template. The red line indicates the donor RNA. Broken red lines are donor RNA cleavage products of RNase H activity. RT is the reverse transcriptase enzyme. Nascent DNA is shown as a dark purple line. Acceptor RNA (green line) invades the donor-nascent DNA hybrid and displaces the DNA onto itself.

5.5 Discussion

Mutated acceptor RNA templates were used to map transfer points during *in vitro* internal strand transfer events in GagPol and Env substrates. A major pause site was earlier mapped to an A residue at position 77 from the 3' end of the donor GagPol template. This corresponds to position 57 from the 3' end of the acceptor GagPol template; 94% of transfer DNAs at 32 minutes of DNA synthesis in the presence of NC contained all three mutations at positions 68, 89 and 105 downstream of the pause site. In conclusion, the nascent DNA transferred either at or within 10 bases downstream of the pause site. The introduction of additional point mutations between bases 57 and 68 can potentially reveal the exact base position. But the number of such mutations is severely limited by the need to conserve the original RNA structure. In addition, the fact that the paused DNA is "chased" away in the presence, but not the absence of acceptor (see Discussions in Chapters 2 and 4) strongly suggests that it is the paused DNA product, and not some other DNA in the mapped region that is transferring. These mapping results corroborate the pause driven mechanism of transfer in highly structured RNA templates like the GagPol. In Chapter 4 it was shown that 5' and 3' truncated and destabilized GagPol acceptor templates promoted strand transfer to the similar levels of efficiency. Taken together, these results suggest a pause induced, DNA dissociation model for transfers in the GagPol templates (Figure 5-9).

The fact that very little pausing is observed on the Env template suggests a pause-independent type mechanism. In Env most transfers were clustered toward the end of the region of homology on the donor. This is consistent with an acceptor

invasion model for transfer (Figure 5-10). The low structure of the Env acceptor would allow rapid binding to the nascent DNA after RT RNase H has degraded the donor. The acceptor then must "zip" up the DNA and displace the donor at a downstream point before DNA synthesis is completed on the donor. Such a scenario would tend to lead to transfers towards the end of the region of homology. This model was supported by dNTP titration assays and mapping results with 1 μM dNTPs. By slowing the RT polymerase, transfer points shifted to upstream locations. In addition, levels of transfer DNA increased significantly, at the expense of the donor-directed DNA.

Structural intricacies of the viral genome were shown to determine the mechanism of transfer events. Even weakly structured genome regions with little potential for pausing can therefore undergo efficient recombination. This is substantiated by the fact that the *env* gene has the most dramatic features of variability. Le *et al* [157, 158] studied seventeen different sequences from various isolates of HIV and showed evidence for a direct correlation between hypervariable clusters and regions with little molecular base pairing. Even though, they found variations throughout the genome, the highest peaks were mainly in the *env* gene. Other significant peaks were found in the *pol* gene (also studied in this report) and one in the *gag* gene but before the *gag-pol* frameshift region. They also propose that single stranded regions of the genome are less constrained in the rate of evolution, more susceptible to mutations and cleavages. Modifications like cleavages would also favor genetic recombination by the force copy-choice method [159]. Although this result supports the idea that recombination is frequent in low structure genome

regions, it is important to note the work was based viral isolates.   In this case

selection for more viable viruses would have occurred and recombinations producing

less fit viruses would have been lost.

Chapter 6   General discussion

The confounding issues facing the development of successful vaccines or drugs to combat the AIDS pandemic are genetic recombination and the high mutation rates of the HIV virus. There are two RNA genomes in the virus that are non-covalently joined together at the 5' end. These two strands are capable of undergoing genetic mutation, mostly due to the low fidelity of RT. Genetic recombination or strand transfer can also occur between the genomes.   Both mutations and recombination have contributed to the emergence of the diverse HIV subtypes or clades [160]. Currently, the HIV strains are classified into three groups: M, O and N of which group M causes over 99% of the world's AIDS cases. Group M is further classified into 10 subtypes or clades: A-J. In patients with chronic infections, the virus causing the initial infection can undergo *in vivo* evolution that generates pools of genetically distinct variants that are referred to as quasispecies [161].   The presence of ancestral sequences within current HIV-1 quasispecies, dated up to 10 years earlier than present ones, has important consequences for in vivo viral evolution, in the pathogenesis and treatment of HIV-1 infection. [162].

Songok *et al.* [163] have investigated *in vivo* evolution by analyzing blood samples from a woman in Kenya who was initially coinfected with subtypes A and D. The samples from 1996 when compared with those from 2002 showed that the *pol* gene from subtype A had recombined with the *env* gene from subtype D resulting in a predominant A/D recombinant form. In another case, intersubtype recombination took place between the initial infecting subtype A and a superinfecting subtype C in an AIDS patient [164]. Robertson and colleagues [165] have reported that eight of

the group M subtypes have recombined to produce genetic hybrids. Recombination between diverse strains of HIV poses problems for eliciting broad range immune responses necessary for a successful vaccine. For example, a vaccine may work against one subtype but may fail against subtype hybrids. This problem is compounded by genetic mutation that results from the highly error prone reverse transcriptase enzyme of HIV. Clinical trials have identified drug resistant mutants to all the FDA approved drugs that are used in therapy of AIDS patients [166, 167]. Recombination between drug resistant mutants has the potential to confer cross resistance to a broad range of drugs, on the resultant hybrid.

A detailed understanding of recombination could target specific proteins or conserved sequences involved in the process, which could identify new approaches for combating the virus. The first area of study in this thesis was the internal strand transfer events that occur in regions of the genome exhibiting different structural intricacies. Many investigators have proposed various models for strand transfer in the viral genome. For example, Chen *et al* [145] proposed an acceptor invasion model for transfer in the TAR region of the HIV genome that is induced by a major pause site and Negroni *et al.*[126] suggested a pause independent model for strand transfer. The experiments conducted in Chapter 2 revealed that strongly structured templates from the U3, *gag-pol* and RRE regions have pause sites that may serve as focal points for transfer. Nevertheless, weakly structured templates from *env* and *pol-vif* regions also transferred efficiently in the absence of major pause sites. These results show that strand transfer events are not constrained by the requirement for strong pause sites and that recombination can occur efficiently from regions with

134

varying thermodynamic stabilities. This is consistent with reports by Jetzt *et al.* [44] that indicate a high rate of recombination throughout the HIV genome. Furthermore, results from Chapter 4 showed that transfer from the pause site on the strong structured GagPol donor is more efficient in the presence of a destabilized GagPol acceptor. This implies that a higher level of recombination may be seen if a highly structured and a relatively low structured RNA genome are copackaged into a virion. This could potentially enhance intersubtype recombination between strains with varying thermodynamic stabilities in homologous regions of the genome. It is difficult to study intersubtype recombination in *in vivo* experiments or animal models. *In vitro* and *ex vivo* assays are good alternatives that can help in identifying 'hotspots' recombination. For example, the envelope protein gp120 is made of five variable regions and four constant regions, designated V1-V5 and C1-C4. Moumen *et al* [174] identified a hotspot for recombination in the C2 region by *in vitro* methods. The same region was found to be a hotspot in *ex vivo* assays conducted by Quinones-Mateu *et al* [175]. Comparison of this hotspot with HIV-1 isolates shows that subgroups A recombine with C or D in the constant portion C2 of the envelope glycoprotein gp120 [178]. This allows for genetic reallocation of the variable regions V1 and V2, relative to regions V3-V5 which helps the virus to escape neutralizing antibodies of the host [179]. In this manner, the regions identified in the above assays can be compared to those in viral isolates from HIV-infected individuals and can help in designing vaccine constructs.

The strongly structured GagPol and the weakly structured Env template were chosen to compare and contrast the mode of transfer in high versus low structured

regions of the genome. Mapping assays and sequencing results from Chapter 5 showed a predominance of transfers from the major pause site at position 77 from the 3' end of the GagPol donor template. Various truncated versions of the GagPol acceptor template were used with a non-variable GagPol donor in Chapter 4 to test for an acceptor invasion mode of transfer. Strand transfer assays showed comparable levels of transfer for 5' and 3' truncated mutant acceptors. This outcome points to a DNA dissociation mechanism rather than acceptor invasion for transfer in the GagPol template. In contrast, the majority of the transfers occurred towards the end of the region of homology in the Env template. These results indicate that the Env acceptor template maybe rapidly associating with the nascent DNA in order to displace it from the donor template before reaching the end of synthesis. The dNTP titration experiments done in Chapter 5 showed that slowing the progression of the RT enzyme gives more time to the Env acceptor to hybridize with the nascent DNA. At low dNTP concentrations (1 μM) transfers were shifted to points located upstream from the previously mapped regions. There was also large drop in the Env donor-directed DNA in the presence of acceptor RNA which not seen in the assays on GagPol. This outcome supports a role for acceptor RNA in inducing transfers in the Env template. Hydroxyurea has been shown to deplete dNTPs in cell cultures and hence slow down the RT enzyme. It was observed that slow polymerase activity produced a higher rate of recombination [176]. These observations have serious implications for the process of drug design, for example, an inhibitor that aims to reduce viral replication by slowing down RT enzyme may not be effective since it would also increase the recombination rate of the virus.

Overall, these analyses suggest that the virus employs multiple methods to overcome problems posed by structural complexities of its genome to successfully complete proviral DNA synthesis. Regions with less structure may undergo a higher frequency of recombination than high structured regions. Eventually, genetic recombination can be viewed as the consequence of the attempts made by the virus to salvage DNA synthesis. Studies on cell cultures have determined that the virus recombines about three times per genome per replication cycle [31]. Rhodes *et al* [116] have reported that HIV-1 recombination can be nearly 10-fold higher than other retroviruses. Many other studies have also reinforced the idea that the virus needs recombination to successfully complete proviral synthesis.

The second area of focus in this study was the NC or nucleocapsid protein. It is a nucleic acid chaperone protein that plays a vital role in many events of the viral life cycle including recombination. Its importance in viral replication has made it an attractive target for antiviral therapy. It is tempting to speculate that anti-NC drugs could inhibit viral infection and or halt viral replication. This is reinforced by studies which show that anti NC drugs that promote ejection of zinc ions from NC zinc fingers can inhibit replication [168]. In addition, molecules have been developed that mimic a part of NC's structure and compete for the recognition of its targets [169]. Some investigators have inactivated human immunodeficiency virus type 1 (HIV-1) with the compound 2,2'-dithiodipyridine. This compound specifically inactivates infectivity of retroviruses by covalently modifying the nucleocapsid zinc finger motifs. Such inactivated viruses have the potential to be successful as whole killed-particle vaccines [170]. Gorelick *et al* [180] used molecular constructs that express

mutant SIV NC to make DNA vaccines. These vaccines generate replication deficient but structurally complete virions that were able to delay the onset of infection in immunized macaques. Another DNA vaccine made with a four amino acid deletion in the second zinc finger of SIV NC was used to immunize 12 macaques [181]. After a challenge infection, all monkeys became infected. Yet, four of them were able to control viral replication and one of the monkeys remained antibody negative.

The ability of NC to accelerate the transfer of minus strand strong stop DNA and to stimulate internal strand exchange *in vitro* has been well documented [100, 101, 171]. NC seems to function by unwinding nucleic acid structures and by annealing complementary strands of RNA or DNA. The results from Chapter 2 showed that transfers in highly structured templates were highly enhanced by NC protein whereas there was only a modest stimulation in weakly structured templates. This illustrates that the NC protein may be more important for recombination in some regions of the viral genome as compared to others. Experiments conducted on zinc finger mutants in Chapter 3 showed that the first zinc finger of NC is crucial for its unwinding property. The role of the second zinc finger of NC is less clear from these experiments. Work by our group and others has confirmed that finger one possesses much of the helix-destabilizing activity of NC while the backbone residues are sufficient for annealing in many experiments [96, 150]. Therefore the role of finger two is unclear, although work in our lab has shown that the unwinding activity of finger one is modestly enhanced if finger two is present [153, 172]. This suggests that finger two may augment the activity of finger one and the backbone amino acids

of NC. It is interesting that simple retroviruses like Moloney murine leukemia virus (MoMLV) possess NC proteins that have only one zinc finger. It has been suggested that HIV may have evolved with two because of the high level of structure near the 5' end of the genome (where dimerization occurs) in comparison to the simple viruses [151]. Future work using mutant NC proteins in which the amino acid residues in the first zinc finger have been serially replaced with those from the second finger could pinpoint the actual residues that direct the helix destabilization activity of NC.

With the emergence of HIV variants that are unresponsive to current treatments and the lack of an effective vaccine it is clear that further studies are warranted. Studies of NC and viral recombination could potentially allow for construction of attenuated or whole killed vaccines that stimulate the immune response without causing disease. Specific compounds that can directly affect recombination or the activity of NC could also be developed. In fact, it has been shown that the anti-cancer drug actinomycin D can inhibit recombination without affecting RT DNA synthesis and RNase H activity [173, 177]. This drug binds to single stranded DNA and interferes with is base pairing capacity. This prevents the transfer of DNA from donor to acceptor RNA. Unfortunately the high level of toxicity of this drug makes it a poor candidate for HIV therapy. However, it remains possible that other less toxic drugs directed against recombination could be developed. A better understanding of the recombination mechanism could certainly aid in this process.

BIBLIOGRAPHY

1.   Barre-Sinoussi, F., et al., *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 868-71.

2.   Gallo, R.C., *The AIDS virus.* Sci Am, 1987. **256**(1): p. 46-56.

3.   UNAIDS, *Global summary of the HIV/AIDS epidemic.* 2003.

4.   Stine, G.J., *Impact of T4 cell depletion.* AIDS update 2001, 2001: p. 146-150.

5.   Stine, G.J., *Opportunisitic infections and cancers associated with HIV disease/AIDS.* AIDS update 2001, 2001: p. 153-179.

6.   UNAIDS, *HIV vaccine trial results are an important step forward in developing an effective vaccine, say WHO and UNAIDS.* Press release, 2003.

7.   VaxGen, *VaxGen announces results of its phase III HIV vaccine trial in Thailand: vaccine fails to meet endpoint.* Press release, 2003.

8.   Stine, G.J., *Preventing the transmission of HIV.* AIDS update 2001, 2001: p. 305-314.

9.   Coffin, J.M., S.H. Hughes, and H.E. Varmus, *Retroviral virions and genomes.* Retroviruses, 1997: p. 27-70.

10.   Hu, W.S. and H.M. Temin, *Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination.* Proc Natl Acad Sci U S A, 1990. **87**(4): p. 1556-60.

11.   You, J.C. and C.S. McHenry, *HIV nucleocapsid protein. Expression in Escherichia coli, purification, and characterization.* J Biol Chem, 1993. **268**(22): p. 16519-27.

12.   Hoxie, J.A., et al., *Biological characterization of a simian immunodeficiency virus-like retrovirus (HTLV-IV): evidence for CD4-associated molecules required for infection.* J Virol, 1988. **62**(8): p. 2557-68.

13.   Sattentau, Q.J., et al., *The human and simian immunodeficiency viruses HIV-1, HIV-2 and SIV interact with similar epitopes on their cellular receptor, the CD4 molecule.* Aids, 1988. **2**(2): p. 101-5.

14.   Sattentau, Q.J. and R.A. Weiss, *The CD4 antigen: physiological ligand and HIV receptor.* Cell, 1988. **52**(5): p. 631-3.

15. Maddon, P.J., et al., *The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain.* Cell, 1986. **47**(3): p. 333-48.

16. Stein, B.S., et al., *pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane.* Cell, 1987. **49**(5): p. 659-68.

17. Haseltine, W.A., et al., *Ordered transcription of RNA tumor virus genomes.* J Mol Biol, 1976. **106**(1): p. 109-31.

18. Haseltine, W.A., J.M. Coffin, and T.C. Hageman, *Structure of products of the Moloney murine leukemia virus endogenous DNA polymerase reaction.* J Virol, 1979. **30**(1): p. 375-83.

19. Shank, P.R., et al., *Mapping unintegrated avian sarcoma virus DNA: termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA.* Cell, 1978. **15**(4): p. 1383-95.

20. Varmus, H.E., et al., *Kinetics of synthesis, structure and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells.* J Mol Biol, 1978. **120**(1): p. 55-82.

21. Darlix, J.L., et al., *Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1.* J Mol Biol, 1990. **216**(3): p. 689-99.

22. Prats, A.C., et al., *Small finger protein of avian and murine retroviruses has nucleic acid annealing activity and positions the replication primer tRNA onto genomic RNA.* Embo J, 1988. **7**(6): p. 1777-83.

23. Temin, H.M. and S. Mizutani, *RNA-dependent DNA polymerase in virions of Rous sarcoma virus.* Nature, 1970. **226**(252): p. 1211-3.

24. Baltimore, D., *RNA-dependent DNA polymerase in virions of RNA tumour viruses.* Nature, 1970. **226**(252): p. 1209-11.

25. Bender, W., et al., *High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures.* J Virol, 1978. **25**(3): p. 888-96.

26. Murti, K.G., M. Bondurant, and A. Tereba, *Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy.* J Virol, 1981. **37**(1): p. 411-19.

27.   Coffin, J.M. and W.A. Haseltine, *Terminal redundancy and the origin of replication of Rous sarcoma virus RNA.* Proc Natl Acad Sci U S A, 1977. **74**(5): p. 1908-12.

28.   Stoll, E., et al., *Avian myeloblastosis virus RNA is terminally redundant: implications for the mechanism of retrovirus replication.* Cell, 1977. **12**(1): p. 57-72.

29.   Taylor, J.M., *An analysis of the role of tRNA species as primers for the transcription into DNA of RNA tumor virus genomes.* Biochim Biophys Acta, 1977. **473**(1): p. 57-71.

30.   Mitra, S.W., et al., *Synthesis of a 600-nucleotide-long plus-strand DNA by virions of Moloney murine leukemia virus.* Proc Natl Acad Sci U S A, 1979. **76**(9): p. 4355-9.

31.   Yu, H., et al., *The nature of human immunodeficiency virus type 1 strand transfers.* J Biol Chem, 1998. **273**(43): p. 28384-91.

32.   Charneau, P., M. Alizon, and F. Clavel, *A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication.* J Virol, 1992. **66**(5): p. 2814-20.

33.   Brown, P.O., et al., *Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein.* Proc Natl Acad Sci U S A, 1989. **86**(8): p. 2525-9.

34.   Roth, M.J., P.L. Schwartzberg, and S.P. Goff, *Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence.* Cell, 1989. **58**(1): p. 47-54.

35.   Le Grice, S.F., et al., *Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase.* Embo J, 1991. **10**(12): p. 3905-11.

36.   Hostomsky, Z., et al., *Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis.* J Virol, 1992. **66**(5): p. 3179-82.

37.   Larder, B.A., et al., *Site-specific mutagenesis of AIDS virus reverse transcriptase.* Nature, 1987. **327**(6124): p. 716-7.

38.   Mizrahi, V., *Analysis of the ribonuclease H activity of HIV-1 reverse transcriptase using RNA.DNA hybrid substrates derived from the gag region of HIV-1.* Biochemistry, 1989. **28**(23): p. 9088-94.

39.   Repaske, R., et al., *Inhibition of RNase H activity and viral replication by single mutations in the 3' region of Moloney murine leukemia virus reverse transcriptase.* J Virol, 1989. **63**(3): p. 1460-4.

40.   Boyer, J.C., K. Bebenek, and T.A. Kunkel, *Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates.* Proc Natl Acad Sci U S A, 1992. **89**(15): p. 6919-23.

41.   Yu, H. and M.F. Goodman, *Comparison of HIV-1 and avian myeloblastosis virus reverse transcriptase fidelity on RNA and DNA templates.* J Biol Chem, 1992. **267**(15): p. 10888-96.

42.   Simon, F., et al., *Identification of a new human immunodeficiency virus type 1 distinct from group M and group O.* Nat Med, 1998. **4**(9): p. 1032-7.

43.   Wain-Hobson, S., *Human immunodeficiency virus type 1 quasispecies in vivo and ex vivo.* Curr Top Microbiol Immunol, 1992. **176**: p. 181-93.

44.   Jetzt, A.E., et al., *High rate of recombination throughout the human immunodeficiency virus type 1 genome.* J Virol, 2000. **74**(3): p. 1234-40.

45.   Zhuang, J., et al., *Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots.* J Virol, 2002. **76**(22): p. 11273-82.

46.   Hu, W.S. and H.M. Temin, *Retroviral recombination and reverse transcription.* Science, 1990. **250**(4985): p. 1227-33.

47.   Katz, R.A. and A.M. Skalka, *Generation of diversity in retroviruses.* Annu Rev Genet, 1990. **24**: p. 409-45.

48.   Temin, H.M., *Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation.* Proc Natl Acad Sci U S A, 1993. **90**(15): p. 6900-3.

49.   Coffin, J.M., *Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses.* J Gen Virol, 1979. **42**(1): p. 1-26.

50.   Zhang, J. and H.M. Temin, *Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication.* Science, 1993. **259**(5092): p. 234-8.

51.   Levy, D.N., et al., *Dynamics of HIV-1 recombination in its natural target cells.* Proc Natl Acad Sci U S A, 2004. **101**(12): p. 4204-9.

52.   Hu, W.S. and H.M. Temin, *Effect of gamma radiation on retroviral recombination.* J Virol, 1992. **66**(7): p. 4457-63.

53. Xu, H. and J.D. Boeke, *High-frequency deletion between homologous sequences during retrotransposition of Ty elements in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 1987. **84**(23): p. 8553-7.

54. DeStefano, J.J., R.A. Bambara, and P.J. Fay, *The mechanism of human immunodeficiency virus reverse transcriptase-catalyzed strand transfer from internal regions of heteropolymeric RNA templates.* J Biol Chem, 1994. **269**(1): p. 161-8.

55. Peliska, J.A. and S.J. Benkovic, *Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase.* Science, 1992. **258**(5085): p. 1112-8.

56. Delviks, K.A. and V.K. Pathak, *Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching.* J Virol, 1999. **73**(10): p. 7923-32.

57. Junghans, R.P., L.R. Boone, and A.M. Skalka, *Retroviral DNA H structures: displacement-assimilation model of recombination.* Cell, 1982. **30**(1): p. 53-62.

58. Tritch, R.J., et al., *Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 gag polyprotein.* J Virol, 1991. **65**(2): p. 922-30.

59. Henderson, L.E., et al., *Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences.* J Virol, 1992. **66**(4): p. 1856-65.

60. Rein, A., L.E. Henderson, and J.G. Levin, *Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication.* Trends Biochem Sci, 1998. **23**(8): p. 297-301.

61. Gorelick, R.J., et al., *Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA.* J Virol, 1990. **64**(7): p. 3207-11.

62. Mely, Y., et al., *Zinc binding to the HIV-1 nucleocapsid protein: a thermodynamic investigation by fluorescence spectroscopy.* Biochemistry, 1996. **35**(16): p. 5175-82.

63. Henderson, L.E., et al., *Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses.* J Biol Chem, 1981. **256**(16): p. 8400-6.

64.   Bess, J.W., Jr., et al., *Tightly bound zinc in human immunodeficiency virus type 1, human T-cell leukemia virus type I, and other retroviruses.* J Virol, 1992. **66**(2): p. 840-7.

65.   Darlix, J.L., et al., *First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses.* J Mol Biol, 1995. **254**(4): p. 523-37.

66.   Chance, M.R., et al., *Extended x-ray absorption fine structure studies of a retrovirus: equine infectious anemia virus cysteine arrays are coordinated to zinc.* Proc Natl Acad Sci U S A, 1992. **89**(21): p. 10041-5.

67.   Mely, Y., et al., *Spatial proximity of the HIV-1 nucleocapsid protein zinc fingers investigated by time-resolved fluorescence and fluorescence resonance energy transfer.* Biochemistry, 1994. **33**(40): p. 12085-91.

68.   Morellet, N., et al., *Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR.* J Mol Biol, 1994. **235**(1): p. 287-301.

69.   Morellet, N., et al., *Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR.* Embo J, 1992. **11**(8): p. 3059-65.

70.   Rice, W.G., et al., *Inhibition of HIV-1 infectivity by zinc-ejecting aromatic C-nitroso compounds.* Nature, 1993. **361**(6411): p. 473-5.

71.   Rice, W.G. and J.A. Turpin, *Virus-encoded Zinc Fingers as Targets for Antiviral Chemotherapy.* Rev Med Virol, 1996. **6**(4): p. 187-199.

72.   Turpin, J.A., et al., *Inhibitors of human immunodeficiency virus type 1 zinc fingers prevent normal processing of gag precursors and result in the release of noninfectious virus particles.* J Virol, 1996. **70**(9): p. 6180-9.

73.   South, T.L., et al., *C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain.* Biochemistry, 1991. **30**(25): p. 6342-9.

74.   Gorelick, R.J., et al., *The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent.* J Virol, 1993. **67**(7): p. 4027-36.

75.   Gorelick, R.J., et al., *Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences.* Virology, 1999. **256**(1): p. 92-104.

76.	South, T.L. and M.F. Summers, *Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC.* Protein Sci, 1993. **2**(1): p. 3-19.

77.	South, T.L., et al., *Zinc fingers and molecular recognition. Structure and nucleic acid binding studies of an HIV zinc finger-like domain.* Biochem Pharmacol, 1990. **40**(1): p. 123-9.

78.	South, T.L., et al., *The nucleocapsid protein isolated from HIV-1 particles binds zinc and forms retroviral-type zinc fingers.* Biochemistry, 1990. **29**(34): p. 7786-9.

79.	De Guzman, R.N., et al., *Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element.* Science, 1998. **279**(5349): p. 384-8.

80.	DeStefano, J.J., *Interaction of human immunodeficiency virus nucleocapsid protein with a structure mimicking a replication intermediate. Effects on stability, reverse transcriptase binding, and strand transfer.* J Biol Chem, 1996. **271**(27): p. 16350-6.

81.	Guo, J., et al., *Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer.* J Virol, 2000. **74**(19): p. 8980-8.

82.	Hargittai, M.R., et al., *HIV-1 nucleocapsid protein zinc finger structures induce tRNA(Lys,3) structural changes but are not critical for primer/template annealing.* J Mol Biol, 2001. **312**(5): p. 985-97.

83.	Schmalzbauer, E., et al., *Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding in vitro.* J Virol, 1996. **70**(2): p. 771-7.

84.	Berthoux, L., C. Pechoux, and J.L. Darlix, *Multiple effects of an anti-human immunodeficiency virus nucleocapsid inhibitor on virus morphology and replication.* J Virol, 1999. **73**(12): p. 10000-9.

85.	Rice, W.G., et al., *Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS.* Science, 1995. **270**(5239): p. 1194-7.

86.	Nagy, K., et al., *Antiviral activity of human immunodeficiency virus type 1 protease inhibitors in a single cycle of infection: evidence for a role of protease in the early phase.* J Virol, 1994. **68**(2): p. 757-65.

87.   Rossio, J.L., et al., *Inactivation of human immunodeficiency virus type 1 infectivity with preservation of conformational and functional integrity of virion surface proteins.* J Virol, 1998. **72**(10): p. 7992-8001.

88.   Karpel, R.L., L.E. Henderson, and S. Oroszlan, *Interactions of retroviral structural proteins with single-stranded nucleic acids.* J Biol Chem, 1987. **262**(11): p. 4961-7.

89.   Lapadat-Tapolsky, M., et al., *Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle.* Nucleic Acids Res, 1993. **21**(4): p. 831-9.

90.   Tsuchihashi, Z. and P.O. Brown, *DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein.* J Virol, 1994. **68**(9): p. 5863-70.

91.   Herschlag, D., *RNA chaperones and the RNA folding problem.* J Biol Chem, 1995. **270**(36): p. 20871-4.

92.   Christiansen, C. and R.L. Baldwin, *Catalysis of DNA reassociation by the Escherichia coli DNA binding protein: A polyamine-dependent reaction.* J Mol Biol, 1977. **115**(3): p. 441-54.

93.   Pontius, B.W. and P. Berg, *Rapid assembly and disassembly of complementary DNA strands through an equilibrium intermediate state mediated by A1 hnRNP protein.* J Biol Chem, 1992. **267**(20): p. 13815-8.

94.   Khan, R. and D.P. Giedroc, *Recombinant human immunodeficiency virus type 1 nucleocapsid (NCp7) protein unwinds tRNA.* J Biol Chem, 1992. **267**(10): p. 6689-95.

95.   Remy, E., et al., *The annealing of tRNA3Lys to human immunodeficiency virus type 1 primer binding site is critically dependent on the NCp7 zinc fingers structure.* J Biol Chem, 1998. **273**(9): p. 4819-22.

96.   De Rocquigny, H., et al., *Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers.* Proc Natl Acad Sci U S A, 1992. **89**(14): p. 6472-6.

97.   Barat, C., et al., *HIV-1 reverse transcriptase specifically interacts with the anticodon domain of its cognate primer tRNA.* Embo J, 1989. **8**(11): p. 3279-85.

98.   Johnson, P.E., et al., *A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription.* Biochemistry, 2000. **39**(31): p. 9084-91.

99.   Rodriguez-Rodriguez, L., et al., *Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro.* J Biol Chem, 1995. **270**(25): p. 15005-11.

100.   You, J.C. and C.S. McHenry, *Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription.* J Biol Chem, 1994. **269**(50): p. 31491-5.

101.   Allain, B., et al., *Transactivation of the minus-strand DNA transfer by nucleocapsid protein during reverse transcription of the retroviral genome.* Embo J, 1994. **13**(4): p. 973-81.

102.   Guo, J., et al., *Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA.* J Virol, 1997. **71**(7): p. 5178-88.

103.   Wu, T., et al., *Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems.* J Virol, 1999. **73**(6): p. 4794-805.

104.   DeStefano, J.J., *Human immunodeficiency virus nucleocapsid protein stimulates strand transfer from internal regions of heteropolymeric RNA templates.* Arch Virol, 1995. **140**(10): p. 1775-89.

105.   Raja, A. and J.J. DeStefano, *Kinetic analysis of the effect of HIV nucleocapsid protein (NCp) on internal strand transfer reactions.* Biochemistry, 1999. **38**(16): p. 5178-84.

106.   Bernacchi, S., et al., *HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence.* J Mol Biol, 2002. **317**(3): p. 385-99.

107.   Wu, W., et al., *Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse transcriptase pausing at a secondary structure near the murine leukemia virus polypurine tract.* J Virol, 1996. **70**(10): p. 7132-42.

108.   Dib-Hajj, F., R. Khan, and D.P. Giedroc, *Retroviral nucleocapsid proteins possess potent nucleic acid strand renaturation activity.* Protein Sci, 1993. **2**(2): p. 231-43.

109.   Amarasinghe, G.K., et al., *NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition.* J Mol Biol, 2000. **301**(2): p. 491-511.

110.   Gorelick, R.J., et al., *Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a "zinc finger-like" protein sequence.* Proc Natl Acad Sci U S A, 1988. **85**(22): p. 8420-4.

111.   Poon, D.T., J. Wu, and A. Aldovini, *Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity.* J Virol, 1996. **70**(10): p. 6607-16.

112.   Poon, D.T., G. Li, and A. Aldovini, *Nucleocapsid and matrix protein contributions to selective human immunodeficiency virus type 1 genomic RNA packaging.* J Virol, 1998. **72**(3): p. 1983-93.

113.   Feng, Y.X., et al., *HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro.* Proc Natl Acad Sci U S A, 1996. **93**(15): p. 7577-81.

114.   Fu, W., R.J. Gorelick, and A. Rein, *Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions.* J Virol, 1994. **68**(8): p. 5013-8.

115.   Fu, W. and A. Rein, *Maturation of dimeric viral RNA of Moloney murine leukemia virus.* J Virol, 1993. **67**(9): p. 5443-9.

116.   Rhodes, T., H. Wargo, and W.S. Hu, *High rates of human immunodeficiency virus type 1 recombination: near-random segregation of markers one kilobase apart in one round of viral replication.* J Virol, 2003. **77**(20): p. 11193-200.

117.   Srinivasan, A., et al., *Generation of hybrid human immunodeficiency virus by homologous recombination.* Proc Natl Acad Sci U S A, 1989. **86**(16): p. 6388-92.

118.   Clavel, F., et al., *Genetic recombination of human immunodeficiency virus.* J Virol, 1989. **63**(3): p. 1455-9.

119.   Roda, R.H., et al., *Strand transfer occurs in retroviruses by a pause initiated two step mechanism.* J Biol Chem, 2002.

120.   Harrison, G.P., et al., *Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5' and 3' of the catalytic site.* Nucleic Acids Res, 1998. **26**(14): p. 3433-42.

121. DeStefano, J.J., et al., *Polymerization and RNase H activities of the reverse transcriptases from avian myeloblastosis, human immunodeficiency, and Moloney murine leukemia viruses are functionally uncoupled.* J Biol Chem, 1991. **266**(12): p. 7423-31.

122. Kati, W.M., et al., *Mechanism and fidelity of HIV reverse transcriptase.* J Biol Chem, 1992. **267**(36): p. 25988-97.

123. Suo, Z. and K.A. Johnson, *Effect of RNA secondary structure on the kinetics of DNA synthesis catalyzed by HIV-1 reverse transcriptase.* Biochemistry, 1997. **36**(41): p. 12459-67.

124. Kim, J.K., et al., *Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1.* J Biol Chem, 1997. **272**(27): p. 16769-77.

125. Wu, W., et al., *Strand transfer mediated by human immunodeficiency virus reverse transcriptase in vitro is promoted by pausing and results in misincorporation.* J Biol Chem, 1995. **270**(1): p. 325-32.

126. Negroni, M. and H. Buc, *Mechanisms of retroviral recombination.* Annu Rev Genet, 2001. **35**: p. 275-302.

127. Adachi, A., et al., *Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone.* J Virol, 1986. **59**(2): p. 284-91.

128. Sambrook, J. and D.W. Russell, *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001. **1**: p. 5.40-5.46.

129. Matzura, O. and A. Wennborg, *RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows.* Comput Appl Biosci, 1996. **12**(3): p. 247-9.

130. Zuker, M. and P. Stiegler, *Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information.* Nucleic Acids Res, 1981. **9**(1): p. 133-48.

131. Zuker, M., *On finding all suboptimal foldings of an RNA molecule.* Science, 1989. **244**(4900): p. 48-52.

132. Jaeger, J.A., D.H. Turner, and M. Zuker, *Improved predictions of secondary structures for RNA.* Proc Natl Acad Sci U S A, 1989. **86**(20): p. 7706-10.

133. Suo, Z. and K.A. Johnson, *RNA secondary structure switching during DNA synthesis catalyzed by HIV-1 reverse transcriptase.* Biochemistry, 1997. **36**(48): p. 14778-85.

134. Le, S.Y., J.H. Chen, and J.V. Maizel, *Thermodynamic stability and statistical significance of potential stem-loop structures situated at the frameshift sites of retroviruses.* Nucleic Acids Res, 1989. **17**(15): p. 6143-52.

135. Mann, D.A., et al., *A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression.* J Mol Biol, 1994. **241**(2): p. 193-207.

136. Malim, M.H., et al., *The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA.* Nature, 1989. **338**(6212): p. 254-7.

137. DeStefano, J.J., et al., *Requirements for strand transfer between internal regions of heteropolymer templates by human immunodeficiency virus reverse transcriptase.* J Virol, 1992. **66**(11): p. 6370-8.

138. Parkin, N.T., M. Chamorro, and H.E. Varmus, *Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo.* J Virol, 1992. **66**(8): p. 5147-51.

139. Kollmus, H., et al., *The sequences of and distance between two cis-acting signals determine the efficiency of ribosomal frameshifting in human immunodeficiency virus type 1 and human T-cell leukemia virus type II in vivo.* J Virol, 1994. **68**(9): p. 6087-91.

140. Cassan, M., et al., *Translational frameshifting at the gag-pol junction of human immunodeficiency virus type 1 is not increased in infected T-lymphoid cells.* J Virol, 1994. **68**(3): p. 1501-8.

141. Dinman, J.D., et al., *The frameshift signal of HIV-1 involves a potential intramolecular triplex RNA structure.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5331-6.

142. Klasens, B.I., et al., *The effect of template RNA structure on elongation by HIV-1 reverse transcriptase.* Biochim Biophys Acta, 1999. **1444**(3): p. 355-70.

143. Negroni, M. and H. Buc, *Copy-choice recombination by reverse transcriptases: reshuffling of genetic markers mediated by RNA chaperones.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6385-90.

144. Negroni, M. and H. Buc, *Retroviral recombination: what drives the switch?* Nat Rev Mol Cell Biol, 2001. **2**(2): p. 151-5.

145. Chen, Y., et al., *Steps of the acceptor invasion mechanism for HIV-1 minus strand strong stop transfer.* J Biol Chem, 2003. **278**(40): p. 38368-75.

146. Summers, M.F., et al., *Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1.* Protein Sci, 1992. **1**(5): p. 563-74.

147. Meric, C. and S.P. Goff, *Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein.* J Virol, 1989. **63**(4): p. 1558-68.

148. De Rocquigny, H., et al., *Two short basic sequences surrounding the zinc finger of nucleocapsid protein NCp10 of Moloney murine leukemia virus are critical for RNA annealing activity.* Nucleic Acids Res, 1993. **21**(4): p. 823-9.

149. Urbaneja, M.A., et al., *Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: elucidation of the structural determinants for function.* J Mol Biol, 1999. **287**(1): p. 59-75.

150. Guo, J., et al., *Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein.* J Virol, 2002. **76**(9): p. 4370-8.

151. Williams, M.C., R.J. Gorelick, and K. Musier-Forsyth, *Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function.* Proc Natl Acad Sci U S A, 2002. **99**(13): p. 8614-9.

152. Carteau, S., R.J. Gorelick, and F.D. Bushman, *Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein.* J Virol, 1999. **73**(8): p. 6670-9.

153. Heath, M.J., et al., *Differing roles of the N- and C-terminal zinc fingers in human immunodeficiency virus nucleocapsid protein-enhanced nucleic acid annealing.* J Biol Chem, 2003. **278**(33): p. 30755-63.

154. Driscoll, M.D. and S.H. Hughes, *Human immunodeficiency virus type 1 nucleocapsid protein can prevent self-priming of minus-strand strong stop DNA by promoting the annealing of short oligonucleotides to hairpin sequences.* J Virol, 2000. **74**(19): p. 8785-92.

155. Reil, H., et al., *A heptanucleotide sequence mediates ribosomal frameshifting in mammalian cells.* J Virol, 1993. **67**(9): p. 5579-84.

156.	Chang, S.Y., et al., *Differential stability of the mRNA secondary structures in the frameshift site of various HIV type 1 viruses.* AIDS Res Hum Retroviruses, 1999. **15**(17): p. 1591-6.

157.	Le, S.Y., et al., *Sequence divergence and open regions of RNA secondary structures in the envelope regions of the 17 human immunodeficiency virus isolates.* Nucleic Acids Res, 1989. **17**(8): p. 3275-88.

158.	Le, S.Y., et al., *Stability of RNA stem-loop structure and distribution of non-random structure in the human immunodeficiency virus (HIV-I).* Nucleic Acids Res, 1988. **16**(11): p. 5153-68.

159.	Parthasarathi, S., et al., *Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications.* J Virol, 1995. **69**(12): p. 7991-8000.

160.	Levy, J.A., *Mysteries of HIV: challenges for therapy and prevention.* Nature, 1988. **333**(6173): p. 519-22.

161.	Delwart, E.L., et al., *Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes.* Science, 1993. **262**(5137): p. 1257-61.

162.	Bello, G., et al., *Co-existence of recent and ancestral nucleotide sequences in viral quasispecies of human immunodeficiency virus type 1 patients.* J Gen Virol, 2004. **85**(Pt 2): p. 399-407.

163.	Songok, E.M., et al., *Active generation and selection for HIV intersubtype A/D recombinant forms in a coinfected patient in Kenya.* AIDS Res Hum Retroviruses, 2004. **20**(2): p. 255-8.

164.	Fang, G., et al., *Recombination following superinfection by HIV-1.* Aids, 2004. **18**(2): p. 153-9.

165.	Robertson, D.L., et al., *Recombination in HIV-1.* Nature, 1995. **374**(6518): p. 124-6.

166.	Schmit, J.C., et al., *Resistance-related mutations in the HIV-1 protease gene of patients treated for 1 year with the protease inhibitor ritonavir (ABT-538).* Aids, 1996. **10**(9): p. 995-9.

167.	Schmit, J.C., et al., *Multiple drug resistance to nucleoside analogues and nonnucleoside reverse transcriptase inhibitors in an efficiently replicating human immunodeficiency virus type 1 patient strain.* J Infect Dis, 1996. **174**(5): p. 962-8.

168.   Huang, M., et al., *Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins.* J Med Chem, 1998. **41**(9): p. 1371-81.

169.   Druillennec, S., et al., *A mimic of HIV-1 nucleocapsid protein impairs reverse transcription and displays antiviral activity.* Proc Natl Acad Sci U S A, 1999. **96**(9): p. 4886-91.

170.   Arthur, L.O., et al., *Chemical inactivation of retroviral infectivity by targeting nucleocapsid protein zinc fingers: a candidate SIV vaccine.* AIDS Res Hum Retroviruses, 1998. p. S311-9.

171.   Peliska, J.A., et al., *Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity.* Biochemistry, 1994. **33**(46): p. 13817-23.

172.   Derebail, S.S., M.J. Heath, and J.J. DeStefano, *Evidence for the differential effects of nucleocapsid protein on strand transfer in various regions of the HIV genome.* J Biol Chem, 2003. **278**(18): p. 15702-12.

173.   Guo, J., et al., *Actinomycin D inhibits human immunodeficiency virus type 1 minus-strand transfer in in vitro and endogenous reverse transcriptase assays.* J Virol, 1998. **72**(8): p. 6716-24.

174.   Moumen, A et al., *Evidence for a mechanism of recombination during reverse transcription dependent on the structure of the acceptor RNA.* J Biol Chem, 2003. **278**(18): p. 15973-82.

175   Quinones-Mateu, M.E., et al., *In vitro subtype recombinants of human immunodeficiency virus type 1: comparison to recent and circulating in vivo recombinant forms.* J.Virol., 2002. **76**(19): p. 9600-9613.

176   Hwang, C.K., Svaroskaia, E.S., Pathak, V.K., *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, 2001. **98**(21): p. 12209-14.

177   Gabbara, S., et al., *Inhibitors of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase.* Biochemistry, 1999. **38**(40): p. 13070-6.

178   Robertson, D.L., et al., *Recombination in HIV-1.* Nature, 1995. **374**(6518): p. 124-6.

179   Ye, Y, et al., *Association of structural changes in the V2 and V3 loops of the gp120 envelope glycoprotein with acquisition of neutralization resistance in a*

simian human immunodeficiency virus passaged in vivo. J Virol, 2000. **74**(24): p. 11955-11962.

180   Gorelick, R.J., et al., *Protection of Macaca nemestrina from disease following pathogenic simian immunodeficiency virus (SIV) challenge: utilization of SIV nuclecapsid mutant DNA vaccines with and without an SIV protein boost.* J Virol, 2000. **74**(24): p. 11935-9.

181   Gorelick, R.J., et al., *Mucosal challenge of Macaca nemestrina with simian immunodeficiency virus (SIV) following SIV nucleocapsid mutant DNA vaccination.* J Med Primatol, 2000. **29**(34): p. 209-19.