NOTE TO USERS

This reproduction is the best copy available.

Role of the NC Protein of Human Immunodeficiency Virus Type 1 in Viral RNA Dimerization and Packaging, as well as in Virus Replication and Stability.

By

Jafar Kafaie, MD.

Division of Experimental Medicine

Faculty of Medicine, McGill University

Montreal, QC, Canada

August 2008

A thesis submitted to McGill University, in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

Your file Votre référence
ISBN: 978-0-494-66600-5
Our file Notre référence
ISBN: 978-0-494-66600-5

I would like to dedicate this thesis to my wife and my daughter.
I would also like to dedicate it to my parents, my sisters and my brother, my late mother in-law.
I appreciate all their love, help and support.

Abstract

In the past three decades, various steps of the human immunodeficiency virus type 1 (HIV-1) life cycle have been thoroughly studied. Many of these steps, such as viral entry, reverse transcription and proteolysis have been targets of antiretroviral therapy. Retroviral genomic RNA (gRNA) dimerization appears essential for viral infectivity and this process appears to be chaperoned by the nucleocapsid (NC) protein of HIV-1. In this dissertation, the role of NC in genome dimerization and other aspects of the viral life cycle have been thoroughly studied. Various positions of the NC protein have been mutated through site-directed mutagenesis and relevant and dispensable positions of NC have been identified through this method. 34 of its 55 residues were mutated, individually or in small groups, in a panel of 40 HIV-1 mutants. It was found that the amino-terminus, the proximal zinc finger, the linker, and the distal zinc finger of NC each contributed roughly equally to efficient HIV-1 gRNA dimerization. The various mutations introduced into NC show the first evidence that gRNA dimerization can be inhibited by: 1) mutations in the N-terminus or the linker of retroviral NC; 2) mutations in the proximal or distal zinc finger of lentiviral NC; 3) mutations in the hydrophobic patch (plateau) or the conserved glycines of the proximal or the distal retroviral zinc finger. Some NC mutations impaired gRNA dimerization more than mutations inactivating the viral protease, indicating that gRNA dimerization may be stimulated by the NC component of the Gag polyprotein (Pr55gag). In the second section of my work, I studied the effect of Pr55gag processing on gRNA dimerization by introducing rate alternating mutants into Pr55gag protein cleavage sites. I showed that Maturation of NCp15 into NCp9 is essential for fast rates of genomic RNA dimerization and maturation of NCp9 into NCp7

has no incidence on genomic RNA dimerization but is essential for viral replication. In order to delineate the amount of viral protease activity needed to produce mature virus 48 hours post transfection, we also studied, by cotransfection studies, the effect of various ratios of wild-type (BH10) and protease-inactive (PR-) plasmids and found that HIV-1 reaches its full genomic RNA dimerization despite 75% unprocessed Pr55gag polyproteins. We have also shown that wild type BH10 plasmid can rescue those mutations in NCp7 protein that have an effect on gRNA dimerization through rescue experiments. Overall, this thesis sheds light on the role of NC in HIV-1 genome dimerization and other aspects of the viral life cycle and identifies the importance of each component of NC during these processes.

Résumé

Au cours des derniers 30 ans, différentes étapes dans la réplication du virus d'immunodéficience humaine de type 1 (VIH-1) ont été soigneusement étudiées. Plusieurs de ces étapes, telles l'entrée du virus, la transcription inverse et la maturation par protéolyse ont été des cibles de la thérapie antivirale. La dimérization de l'ARN génomique rétroviral (gRNA) semble essentielle pour la réplication virale et ce processus semble contrôlé par la protéine de nucléocapside (NC) du VIH-1. Dans cette thèse, le rôle de NC dans la dimérization génomique, ainsi que d'autres aspects du cycle de vie viral ont été soigneusement étudiés. Différents résidus de NC ont été mutés par mutagenèse ciblée et des résidus essentiels et inutiles ont été identifiés à l'aide de cette méthode. 34 des 55 résidus ont été mutés, individuellement ou en groupe, à l'aide d'un portefeuille de 40 mutants. Nous avons trouvé que le bout amino, le doigt zinc proximal, le lien, et le doigt zinc distal de NC contribue chacun à peu près également à une dimérization efficace. Les diverses mutations introduites dans NC montrent pour la première fois que la dimérization peut être inhibée par des mutations: 1) dans le bout amino ou le lien de NC rétroviral; 2) dans le doigt zinc proximal du NC lentiviral; 3) dans la plaque hydrophobique or les glycines conservées du doigt zinc rétroviral proximal ou distal. Quelques mutations dans NC inhibent la dimérization du gRNA plus que des mutations inactivant la protéase virale, indiquant que la dimérization du gRNA peut être stimulée par la composante NC de la polyprotéine Gag (Pr55gag). Dans la 2e partie de cette thèse, j'ai étudié l'effet de la maturation de Pr55gag sur la dimérization du gRNA en introduisant, dans les sites de clivage, des mutations qui altèrent le taux ou la vitesse de maturation de Pr55gag. J'ai montré que la maturation de NCp15 en NCp9 est essentielle

pour une dimérization rapide du gRNA, et que la maturation de NCp9 en NCp7 n'a aucune incidence sur la dimérization du gRNA mais est essentielle pour la réplication virale. De manière à délimiter la quantité de protéase virale nécessaire pour produire des virus qui sont matures après une transfection de 48 h, nous avons aussi étudié, par cotransfection, l'effet de différentes proportions de plasmides sauvages (BH10) et protease inactive (PR-). Nous avons trouvé que HIV-1 réussi à dimériser complètement son gRNA malgré que Pr55gag soit 75 % non maturé. Nous avons aussi montré que le plasmide sauvage peut annuler l'effet des mutations NC introduite dans un 2e plasmide utilisé en contransfection. En tout, cette thèse illumine le rôle de NC dans la dimérization du gRNA et certains autres aspects du cycle de vie viral, et identifie l'importance de chaque partie de NC dans ces processus.

Preface

This dissertation is written in the form of manuscript-based thesis according to the "guidelines for thesis preparation" from the "Faculty of Graduate studies and Research at McGill University":

"As an alternative to the traditional thesis format, the thesis can consist of a collection of papers of which the student is an author or co-author. These papers must have a cohesive, unitary character making them a report of a single program of research. The structure for the manuscript-based thesis must conform to the following: Candidates have the option of including, as part of the thesis, the text of one or more papers submitted, or to be submitted, for publication, or the clearly-duplicated text (not the reprints) of one or more published papers. These texts must conform to the "Guidelines for Thesis Preparation" with respect to font size, line spacing and margin sizes and must be bound together as an integral part of the thesis (reprints of published papers can be included in the appendices at the end of the thesis)."

This dissertation contains 4 chapters. Chapter 1 is the background and introduction to the topic. Chapter 2 and 3 are based on the manuscripts listed below. Chapter 4 is the conclusion chapter where the contributions to original knowledge have been discussed.

The manuscripts presented in this thesis are the following:

Chapter 2:

Kafaie J, Song R, Abrahamyan L, Mouland AJ, Laughrea M. **Mapping of Nucleocapsid Residues Important for HIV-1 Genomic RNA Dimerization and Packaging**.Virology. 2008 Jun 5;375(2):592-610. Epub 2008 Mar 17.

This manuscript has been published in Virology. Song, R. has contributed to discussions. Abrahamyan, L. has performed the western blots and the laboratory facilities of Dr. A. Mouland were used for some of the experiments presented.

Chapter 3:

Jafar Kafaie, Marjan Dolatshahi, Lara Ajamian, Rujun Song, Andrew J. Mouland, Isabelle Rouiller, Michael Laughrea. **Role of Capsid Sequence and Immature Nucleocapsid Proteins p9 and p15 in Human Immunodeficiency Virus type 1 Genomic RNA Dimerization**. Virology. 2009 Mar 1; 385 (1): 233–244. Epub 2008 Dec 13.

This manuscript contains Cryo-electron-microscopy studies that are not included in this thesis. For the sections included in my thesis, Lara Ajamian has performed the western blots and Dr. Mouland provided me with laboratory space.

Other work not presented in this thesis:

I have also contributed to two other manuscripts not included in this thesis. My contribution as co-author was on a level of technical assistance and scientific support:

1) Song R, Kafaie J, Laughrea M. **Role of the 5' TAR Stem-Loop and the U5-AUG Duplex in Dimerization of HIV-1 Genomic RNA**. Biochemistry. 2008 Mar 11;47(10):3283-93. Epub 2008 Feb 16.

2) Song R, Kafaie J, Yang L, Laughrea M. **HIV-1 viral RNA is Selected in the Form of Monomers that Dimerize in a Three-Step Protease-Dependent Process; the DIS of Stem-Loop 1 Initiates viral RNA Dimerization**. J Mol Biol. 2007 Aug 24;371(4):1084-98. Epub 2007 Jun 9.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Michael Laughrea for his support, leadership and guidance during the past years.

I would like to thank the director of experimental medicine, Dr. Hugh P.J. Bennett as well as Mrs. Dominique Besso and Mrs. Marylin Linhares for their help and support during my studies.

I would like to acknowledge Dr. Beitel for all the help.

I would like to thank Dr. Heather D. Durham for her great help, support and understanding.

I would also like to thank my academic advisor Dr. Volker Blank for his generous helps and supports.

I like to show my appreciation to the members of my graduate committee: Dr. Chantal Autexier, Dr. Lea Brakier-Gingiras and Dr. Lawrence Kleiman who have all been very helpful and kind to me.

I would like to thank Dr. Andrew Mouland, Levon Abrahamyan and Lara Ajamian for collaborating on this project.

I would like to acknowledge Dr. Mark Wainberg, Gabriella Di Pancrazio, Rita Notte and Bianca Di Iorio for the provided facilities and the administrative assistance.

I would like to thank my friends Moutih Rafei, Yousef Alhaji, Gity Mousavi and Maryam Ehteshami for their help.

I would like to thank my colleges in our lab: Louis Jette, Ni Shen, and Rujun Song for their help and collaboration during these years.

I would like to acknowledge the technicians at LDI who were very helpful to me: Cesar, Maureen, Bonnie and Carlos.

Finally, I would also like to warmly thank Dr. Mark Trifiro who was not only my physician, but also a great source of inspiration for me.

Table of contents

List of Figures and Tables

Chapter 1

Chapter 2

List of Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
BMH	branched multiple hairpin
CDC	Centers for Disease Control
C	Cysteine
CA	Capsid
CRF	circulating recombinant form
DIS	dimer initiation site
gRNA	genomic RNA
HIV	Human Immunodeficiency Virus
IN	Integrase
IRES	Internal Ribosome Entry Segment
LDI	long distance interaction
M	membrane-binding
MA	membrane-associated
MHR	major homology region
MoMuLV	Moloney murine leukemia virus
NC	nucleocapsid
NNRTI	non-nucleoside-analogue RT inhibitor
NRTI	nucleoside-analogue RT inhibitor
PBS	Primer Binding Site
PI	protease inhibitor
PIC	pre-integration complex

PPT	polypurine tract
PR	protease enzyme
PR^-	protease inactive
Pr55gag	Gag precursor polyprotein
Pr160gag-pol	Gag-Pol precursor polyprotein
RNP	ribonucleoprotein
RRE	Rev response element
RT	reverse transcriptase
SIV	Simian Immunodeficiency Virus
SL	stem loop
SU	subunit
TM	transmembrane
UTR	untranslated region
WT	Wild Type

Preface to Introduction

In this chapter, I will first include a brief section on the history, epidemiology and pathogenesis of AIDS and HIV-1. This section will be followed by a review of HIV-1 virion structure and genome arrangement. I will then review our current understanding of the role of NC protein in HIV-1 life cycle. In this sub-section, first I will explain the structure of NC protein and then its role in HIV-1 viral life cycle will be explained in two major categories: early and late life cycle. HIV-1 genomic RNA dimerization is the last topic in this chapter where I will provide a description of HIV-1 genomic RNA dimerization and review the role of NC protein during this process.

Chapter 1

Background and Introduction

1.1. Introduction

The Human Immunodeficiency Virus (HIV) is the causative agent of Acquired Immunodeficiency Syndrome (AIDS) which has been responsible for high rates of morbidity and mortality throughout the world ever since the onset of the epidemic in the early eighties (1). HIV belongs to the *lentiviridae* genus of retroviruses. Retroviruses have a single stranded RNA genome from which a double stranded DNA copy can be made and integrated into the host genome (1,2). The term *lentiviridae* refers to the long time it takes for disease development after the initial infection. The HIV virus is a highly diverse quasispecies with multiple types and subtypes (2,3). Two main types of HIV have thus far been identified namely HIV-1 and HIV-2. The former is the predominant form while the latter shares more genomic similarities with Simian Immunodeficiency Virus (SIV) (4). The pandemic causing HIV-1 type can be further grouped into three separate categories: M (main), O (outlier) and N (non-M/non-O) (5). The M group consists of 11 different clades or subtypes (Figure 1.1). When genetic recombination occurs between two subtypes, the resulting virus is referred to as a circulating recombinant form (CRF). Individuals infected with HIV could be harboring either of the subtypes or CRF strains. While most of the HIV-positive population is infected with HIV-1, HIV-2 infections can be found in West African nations (1). Subtype B is the prevalent strain in Europe and North America (2).

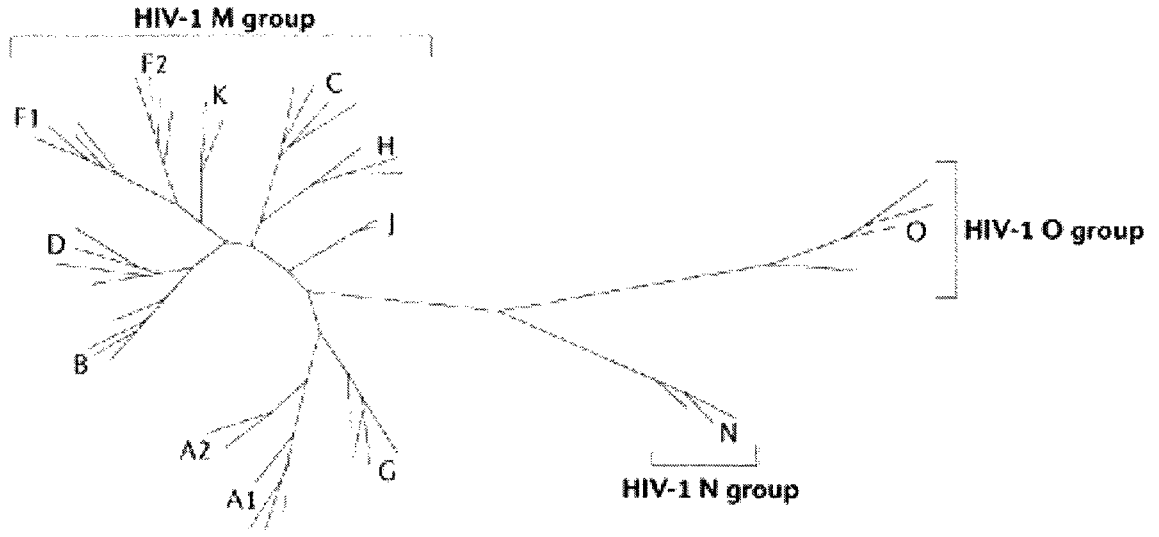


FIGURE 1.1. Phylogenetic classification of HIV-1 (adapted from (5)).

1.1.2. A brief history of HIV and recent statistics

In the summer of 1981, the Center for Disease Control (CDC) reported five cases of rare pulmonary infection with *Pneumocystis carinii* in young homosexual men (6). Due to immune deterioration, the patients experienced severe infections with rare opportunistic organisms and developed malignant transformation. At first it was assumed that Acquired Immunodeficiency Syndrome (AIDS), a name used first by CDC in 1982, is due to immune overwhelming in some high risk populations. But in late 1982, there was enough epidemiologic evidence to suggest that an infectious agent, transferred by bodily fluids and blood products was responsible for the infection outbreaks (7). Immunological investigations paved the way towards the discovery of the true cause of AIDS (1). Early research mainly focused on the rapid decline of circulating CD4 T helper cells. It was observed that once the number of circulating CD4 T cells fall below 200 cell/mm^3 in peripheral blood, the patients started to develop opportunistic infections and malignancies (8). Interestingly enough, the same cell subgroup was also the target of the newly

described human retrovirus, HTLV-1 by Robert Gallo in 1980 (9). But instead of decreasing the number of CD4 T helper cells in the periphery, HTLV-1 transforms these cells leading to Adult T-cell Leukemia (10). In 1984 the same investigator described a causal link between a new retrovirus and AIDS. Finally in 1986, the pathogen, thus far known by different names, was recommended to be called the Human Immunodeficiency Virus (HIV) (11). The history of HIV research is full of hopeful achievements and bitter disappointments. In spite of considerable effort to control the global pandemic, it still continues to expand throughout the world, with especially devastating effects in sub-Saharan Africa (1).

Since its appearance 27 years ago, 65 million people have been infected with this retrovirus, and more than 25 million people have already died of AIDS (1). In spite of its decline in developed countries, it is still in its high spreading wave through developing countries, where everyday 1600 new born babies continue to acquire the virus from their mother during the labor. More than 12 million children have lost their parents to AIDS. This number is estimated to rise to 25 million by 2010 (12). In highly affected countries, women in the 15 to 24 years age group represent 60 percent of the infected population. In 2006, 2.8 million people died because of AIDS and 4.1 million were newly infected (1). Recent years have seen a steady increase in our understanding of the pathophysiology of the disease and the structure and life cycle of the HIV virus. As a result, a large number of antiretroviral agents have been developed and their correct administration can result in prolonging the life of the infected individual. However, vaccine development efforts have not been as successful and the eradication of AIDS is still far from reach.

1.1.3. Transmission of HIV

HIV can be transmitted through several different pathways. The main route of transmission is through unprotected sex with an infected individual. This can be vaginal, anal or oral while the latter is of lower risk unless there is a cut in the partner's mouth (13). Sharing of syringes and needles between infected and non-infected peoples is a common mode of transmission especially among injection drug users (13). Infection can also occur through contact with infected, blood and blood products such as unscreened transfusion blood. Mother to child transmission can occur in three ways: babies born to infected mothers can get the virus during prenatal (through mother-fetus blood exchange inside the uterus), natal (through direct contact with maternal blood during labor) and post-natal period of life (through breast feeding) (14). Finally, in a hospital setting, health care professionals must also take special precaution so as to avoid contact with blood containing HIV. Because semen and vaginal fluids contain high amount of viruses, and due to the possibility of damage to the mucosal membrane during sexual intercourse, this route still remains the most prominent way of HIV transmission (13). As such, the proper use of condoms still remains the best means of HIV prevention. Finally, recent literature has identified that male circumcision can significantly reduce the rate transmission of HIV (15). In some studies, a 50% reduction in transmission was observed suggesting that this practice may play a key role in controlling the spread of HIV (8,12,16).

1.1.4. Pathogenesis of HIV infection

AIDS refers to a collection of infections, malignancies and symptoms that result from the weakening of the immune system due to HIV infection. During normal immune system function, infection results in antigen presentation, which leads to CD4 T cell (or helper T cell) activation. This leads to the secretion of cytokines and other signaling molecules which further activate

other components of the immune system such as B cells, natural killer cells and CD8 T cells. The activated immune cells are in the front line of fighting infection (17). HIV infects CD4 T cells specifically and this handicaps the orchestration of the immune response. There are approximately 800-1200 CD4 T cells per mm^3 of peripheral blood in a healthy person. HIV propagation in the body however results in the slow destruction of CD4 T cells over the course of several years (typically about 10). At its height, virus replication can lead to a drop in the CD4 T cell count as low as 200 cells per mm^3 of peripheral blood or even lower. As such, when the immune system is suppressed, the infected individual becomes highly susceptible to developing malignancies and contracting various opportunistic infections that would not normally cause harm to a healthy individual (18). Examples of such infections include Candidiasis, Kaposi's sarcoma, cryptosporidiosis, pulmonary tuberculosis, and herpes zoster. Symptoms during early HIV infection resemble acute viral infection and include fever, weight loss, fatigue, headache, recurrent diarrhea, etc (12). If access to antiretroviral is provided, the decline in CD4 T cell count can be slowed down and the onset of AIDS can be delayed in most individuals. Overall, HIV infection has proven to be a complex process affecting various cell types such as macrophages, dendritic cells and Langerhan cells. In addition HIV infection of the central nervous system also leads to dementia and other neurological dysfunctions. Despite significant progress in understanding HIV infection, many questions still remain unanswered regarding HIV-mediated immune depletion and the host's immune response (2).

1.1.5. HIV treatment

In the absence of a cure or a successful immunization strategy against HIV, antiretroviral therapy currently represents the most promising development in the fight against AIDS. Antiretroviral

therapy includes the co-administration of multiple drugs during cocktail therapy. Established regimens generally include the administration of nucleoside-analogue RT inhibitors (NRTIs), non-nucleoside-analogue RT inhibitors (NNRTIs) or protease inhibitors (PIs)(2). NRTIs constitute the major class of antiretrovirals with drugs such as zidovudine, stavudine and tenofovir (19). While NNRTI drug nevirapine appears to promote the prevention of mother to child transmission of HIV, the efficacy of most NNRTI compounds is undermined due to the development of cross-resistance. A novel class of NNRTI compounds named Etravirine has recently been approved for clinical use. Unlike other NNRTIs, the molecular flexibility of this drug increases the resistance barrier and allows for its administration to treatment-experienced patients (20). Antiretrovirals belonging to novel classes of inhibitors have recently been approved for clinical usage. These include entry inhibitors such as Maraviroc and Enfuvirtide as well as the integrase inhibitor Raltegravir (21). Maraviroc is a CCR5 antagonist which blocks the interaction of CCR5 co-receptor with gp120 of HIV-1 (22). Enfuvirtide on the other hand, is a small peptide molecule which can bind to gp41 and prevent the fusion of the virion with the cell plasma membrane (23). Together, there are currently around 30 different antivirals that are administered to HIV-positive individuals. Prophylactic and therapeutic vaccine strategies have also been thoroughly investigated albeit they have proven unsuccessful to date. Such strategies include the development of DNA vaccines, recombinant protein vaccines as well as recombinant or attenuated viral vaccines (2,24). Another area of research and discovery where our knowledge had greatly expanded in the past few years is the discovery that host restriction factors such as APOBEC and TRIM5alpha play a role in controlling viral infection (25,26). The therapeutic relevance of these factors is currently under intense investigations.

To summarize, antiretrovirals are the leading weapon for combating the disease progression. Their combination has had a significant impact on improving morbidity and mortality in infected patients as well as the prevention of mother to child transmission. However, problems of adherence, toxicity and the eventual development of resistance are significant obstacles in the treatment of HIV. Because of this, the development of preventive measures and compounds inhibiting novel HIV targets could play a key role in controlling the HIV pandemic.

1.1.6. Virion structure and genome

The HIV-1 virion consists of an envelope where viral proteins gp41 and gp120 are inserted into a cell-derived lipid bilayer (Figure 1.2.). This envelope contains the conical capsid formed from viral capsid proteins (p24). Ribonucleoprotein (RNP) is inside the core, which consists of two copies of the RNA genome as well as the nucleocapsid protein (p7). Other viral enzymes such as reverse transcriptase (p51/p66), protease (p11) and integrase (p32) are also associated with RNP. The coordinated presence of all the viral components allow for the formation of an infectious virus (2).

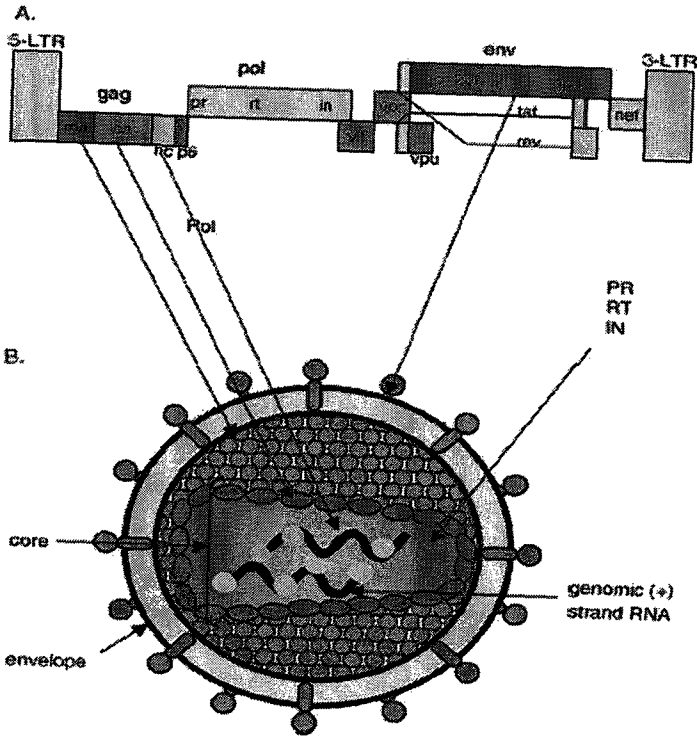


FIGURE 1.2. HIV-1 virion structure and genome (adapted from (27)).

The packaged genome of the HIV virus consists of two copies of 9.2 Kb positive, single-stranded RNA molecules. The HIV genome consists of three major structural genes, Pr55*gag*, Pr160*gagpol* and *env* (28,29). Pr55gag and Pr160gagpol are initially expressed as large precursor polyproteins by free ribosomes, and get subsequently cleaved by the viral protease to yield functioning viral proteins. The *Env* gene is translated by ribosomes of the RER and, after being post translationally modified by multiple glycosylation and cleavage by a cellular enzyme, gets transferred to the cell membrane through the Golgi system of the cell. In addition to the major genes which code for structural and enzymatic proteins, the HIV-1 genome also contains six auxiliary genes named *tat, rev, nef, vif, vpr* and *vpu,* the functions of which will be discussed in the following paragraphs. The genetic organization of the HIV-1 genome is depicted in figure 1.3. In this section, the functions of each of the viral gene products are individually discussed.

1.1.7. The *gag* gene products

Pr55gag is translated from full length RNA on free ribosomes in the cytoplasm of infected cell. After its cleavage during viral maturation, the Gag polyprotein ultimately gives rise to several structural proteins such as MA, CA, NC (30). Even in the absence of any protease-mediated processing, the Gag precursor is sufficient for virus assembly in the host cell. In fact, when all other viral proteins are absent, Gag alone is sufficient for the release of virus-like particles. Gag contains the elements required for targeting Gag and Gag-Pol to the host cell plasma membrane. It also directs the incorporation of viral genome and other necessary molecules into the budding viral particle (2).

MA. One of the proteins expressed by the *gag* gene is the matrix or, membrane-associated protein (MA). MA lies immediately under the viral membrane and forms the viral matrix. It is part of the N-terminus of the Gag polyprotein and plays several roles during the viral replication cycle. The membrane-binding (or M) domain of MA located at the N-terminus of the protein is myristylated and serves to target Gag and Gag-Pol to the host cell plasma membrane (31-33). Mutations within this region have been shown to severely compromise the stability of membrane binding and particle assembly (34,35). It has also been demonstrated that MA is part of the pre-integration complex (PIC) and that the presence of a nucleus localization signal at the C-terminus of MA allows for this viral protein to play a critical role in translocating PIC into the nucleus (36-38). MA has also been implicated for the proper incorporation of envelope glycoproteins into newly-formed virions as well as early steps of infection prior to reverse transcription (39-41).

CA. Also expressed by the *gag* gene, CA represents the capsid protein that forms the core of HIV-1. CA plays an essential role during virion assembly and release (42). The C-terminal segment of CA contains the conserved sequence MHR (major homology region) which allows

for CA dimerization. This region also plays a role in Gag oligomeization during viral assembly and release. On the other hand, the N-terminal region of CA is not necessary for particle assembly but plays an important role in the formation of the mature viral core and is hence essential for the infectivity of the virus (43-46).

NC. In brief, the nucleocapsid (NC) protein contains two zing finger domains that allow for nucleic acid binding. NC plays multiple critical roles at various stages of the viral cycle. These roles include core formation, genome dimerization and encapsidation, primer tRNA positioning and unwinding, stimulation of reverse transcription and interactions with the Vpr protein (2,47). The functions of NC are thoroughly discussed in section 1.2.1.

p6. Also known as $p6^{gag}$, is located at the 3' end of the Gag precursor protein but is not present in the Gag-Pol precursor due to a frameshift during translation. P6 appears to promote the incorporation of Pol and Vpr in the viral particles and contributes to efficient budding of the virion by facilitating the release of viral particles. This protein also promotes viral maturation by down-regulating viral protease activity in the budding virion (48,49).

p1 and p2. Small spacer peptides 1 and 2 (p1 and p2) are generated by the proteolytic cleavage of Gag. While their location in the gene remains conserved, the length and sequence of these proteins are variable. p2 appears to play a critical role in regulating the sequential processing of Gag. When p2 is deleted from the viral genome, the viral morphology is disrupted, resulting in loss of infectivity. The role of p1 is not as clearly defined. However, this protein appears to play a role in viral RNA encapsidation (2,50,51).

1.1.8. The *pol* gene products

The *pol* gene codes for viral peptides with enzymatic activity. The Gag-Pol precursor protein is expressed with Gag expression. This is due to the fact that the Gag-Pol is synthesized via a -1 translational frameshift during translation of the p1-doding part of the viral RNA. Due to the low frequency of the frameshift, which allows for Gag-Pol expression, the ration of Gag-Pol to Gag alone is 1:20 (52,53). The Gag-Pol precursor is incorporated into the viral particle upon assembly. In addition to containing essential enzymes for viral replication, the Gag-Pol precursor can itself enhance viral particle stability whereby the Gag region of Gag-Pol targets the precursor to the virion (2).

PR. The protease enzyme of HIV (PR) is an essential component of the viral life cycle (54,55). HIV-1 PR belongs to the aspartic family of PRs due to the presence of the DTG sequence in its active site (56). A functional PR is a homodimer consisting of two 11-kDa monomers. The monomers are symmetrically associated through their C and N termini and together contribute to the formation of the substrate binding cleft. PR functions by recognizing specific proteolytic sequences in its substrate which allow for the specific proteolytic cleavage of substrates (57-59).

RT. The reverse transcriptase (RT) enzyme of HIV is responsible for converting the viral genomic RNA into double-stranded proviral DNA. RT functions as a heterodimeric molecule consisting of the p66 subunit associated with the p51 subunit. The latter has the same sequence as p66 but is missing the C-terminal region. While p51 has a structural role in RT function, the p66 subunit contains the enzymatic active sites of RT consisting of the polymerase active site and the RNAse H active site. As such, RT has three catalytic activities: RNA-dependent-DNA polymerization, DNA-dependent-DNA polymerization and RNA degradation during reverse transcription (60,61).

IN. Integrase (IN) is the third major protein expressed and it forms the C-terminus of Pol. This enzyme functions as a tetratmer (dimer of dimers) and is responsible for the integration of the proviral DNA into the host cell genome (62,63). HIV-1 genome integration can be site-specific through the LTR region of the viral DNA and non-specific in terms of the sites of integration in the host chromosome. In addition to integration, knock-out studies suggest that IN can play a role in assembly and maturation of viral particles (64,65).

1.1.9. The *env* gene products

The singly spliced RNA coding for *env* originally gives rise to the precursor polyprotein gp160 which is transported to the plasma membrane for incorporation into newly formed viral particles. After maturation and proteolytic cleavage, gp160 gives rise to TM and SU which then undergoes further glycosylation and other modifications and are subsequently inserted into the virion membrane (2).

SU. The viral gp120 envelope glycoprotein is known as the external subunit (SU) protein. During target cell recognition, SU binds to host CD4 receptors as well as coreceptors (CCR5 or CXCR4 depending on viral tropism). This interaction triggers conformational changes that lead to the exposure of TM and allow for viral entry (66,67).

TM. The transmembrane protein TM also known as gp41 is expressed at the C terminal of Env. This protein serves as the anchor that mediates the pH-independent fusion of viral membrane with host cell plasma membrane during viral entry (2,68).

1.1.10. Accessory Proteins

In addition to the major gene products described above, accessory proteins are expressed at different time points during the viral life cycle. Depending on whether or not the expression of

these proteins is regulated by Rev, the accessory proteins are classified as early or late proteins. Tat, Rev and Nef are considered "early" proteins while Vpr, Vif and Vpu are "late" proteins (2).

Rev. Located in the host cell nucleus, the Rev protein is expressed early during the viral life cycle and allows for the temporal regulation of viral gene expression (28). It does so by allowing the translocation of unspliced or singly spliced viral RNA from the nucleus into cell cytoplasm where protein translation can occur. This mRNA translocation is mediated by the binding of Rev to the Rev response element (RRE) on mRNA. Ultimately, the accumulation of Rev and mRNAs in the cytoplasm allow for the regulation of "late" protein expression (69-71).

Tat. This small viral protein acts as a transactivator of gene expression by enhancing LTR-mediated transcription of viral genes. Tat binds to the TAR region of viral RNA and increases both transcription initiation and RNA polymerase II processivity (71-73).

Nef. The exact role of this protein during the viral life cycle still remains to be fully elucidated. Nef, which is expressed throughout the viral life cycle, appears to modify T cell activation and hence promote viral gene expression. Nef also down-regulates CD4 (74) and MHCI (75) expression in the infected T cell and can hence contribute to viral pathogenesis (76). This protein has also been implicated in inducing apoptosis in HIV specific bystander CD4 T cells (77).

Vif. As a "late" protein, it appears that Vif can enhance viral infectivity by interacting with APOBEC3G which is host restriction factor. APOBEC3G action results in G to A mutations in the viral genome, pushing the virus towards error catastrophe. Vif inhibits the action of APOBEC3G by mediating the degradation of this restriction factor in proteosome-dependent manner (78,79). This protein also plays a role in proper viral core formation and virion transport to the nucleus through cell microfilament machinery (80,81).

Vpu. Expressed as a "late" protein, Vpu helps with the down-regulation of CD4 receptors and promotes the transport and maturation of the Env protein (82). Vpu also plays a role in enhancing viral release from the host cell by forming ion channels in the cell membrane (83).

Vpr. Is expressed early during the viral life cycle and is located to the host cell nucleus. Through its interaction with p6, high levels of Vpr get incorporated into the formed viral particle (84). This subsequently allows for Vpr to transport the PIC towards the nucleus upon infection (85). Vpr also plays a critical role in mediated cell cycle arrest in the G_2 phase and hence actively contributes to viral reproduction and infectivity (86).

1.1.11. HIV-1 viral replication cycle

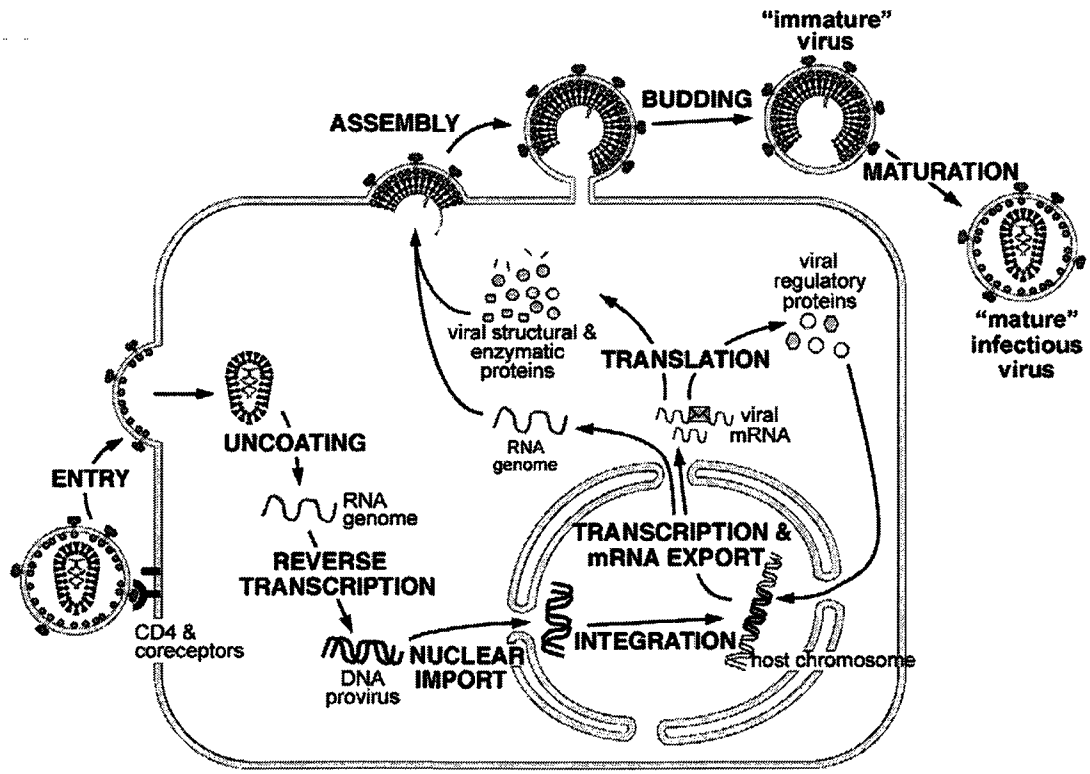


FIGURE 1.3. Schematic representation of HIV life cycle (adapted from(87)).

HIV entry. HIV-1 infection is initiated with the entry of the viral particle into susceptible host cells. Viral envelope glycoprotein gp120 recognizes and binds to CD4 T cell surface receptor, as well as chemokine receptors CXCR4 or CCR5, which serve as co-receptors during viral infection (88). CXCR4-tropic (T-tropic) viral particles, which generally appear at the later stages of HIV infection, can infect T cells and is thought to be syncytium inducing. CCR5-tropic (M-tropic) viruses on the other hand, can infect macrophages and do not induce syncytium formation in T cells (89-92). CD4 receptor can serve as a docking surface for the viral particle and promote virus-coreceptor interactions. The main determinant of tropism of HIV virus is the highly variable V3 loop of the gp120 glycoprotein (93-95). When CD4 binds to gp120, a conformational change is induced that allows for the V3 loop to come in contact with the appropriate coreceptor. This interaction triggers further conformational changes that lead to the exposure of gp41, the other viral glycoprotein essential for viral fusion. Once gp41 is exposed, it inserts itself into the cell target membrane and allows for the fusion of viral and plasma membrane in a pH-independent manner (96-98).

Reverse Transcription. Once the viral core is in the cytoplasm, the reverse transcriptase (RT) enzyme converts the viral RNA genome into double-stranded proviral DNA. This process is presented in figure 1.4. The following describes the step-by-step mechanism of HIV-1 reverse transcription (2,19,99,100).

1. A host cell $tRNA^{Lys3}$ molecule initially serves as the primer for the synthesis of the minus-strand DNA. With the assistance of NC, it is annealed to the Primer Binding Site (PBS) near the 5' end of the RNA strand. Synthesis continues until the 5' end of the template. This product is referred to as minus strand strong stop DNA.

2. The RT enzyme contains an RNase H active site, which degrades the RNA template part of the RNA/ strong stop DNA duplex as this strand is being synthesized. This allows for the strand transfer of the DNA to the 3' end of the RNA.
3. Once the minus strand strong stop is fully synthesized, it is transferred to the 3' region of the RNA genome where sequence complementarity between the R region of both sequences allows for their annealing.
4. Elongation of the minus strand DNA can continue simultaneously with RNA degradation.
5. However, the polypurine (PPT) tract remains uncleaved and it serves as the primer for the synthesis of the positive DNA strand. Plus strand DNA synthesis continues until 12 nucleotides are incorporated. At this point the enzyme dissociates and reassociates in a manner that allows for PPT RNA degradation. RT then rebinds once again in the orientation that allows for DNA polymerization.
6. Upon reaching the tRNA region, plus strand synthesis is stopped and the tRNA primer is degraded allowing for the second strand transfer event.
7. A second strand transfer occurs when PBS region on the plus strand DNA is annealed to the PBS sequence on the minus strand DNA. This leads to the circularization of the complex.
8. DNA synthesis is then completed on both strands. Duplication of the U5 and U3 regions leads to the generation of a blunt-ended proviral DNA.

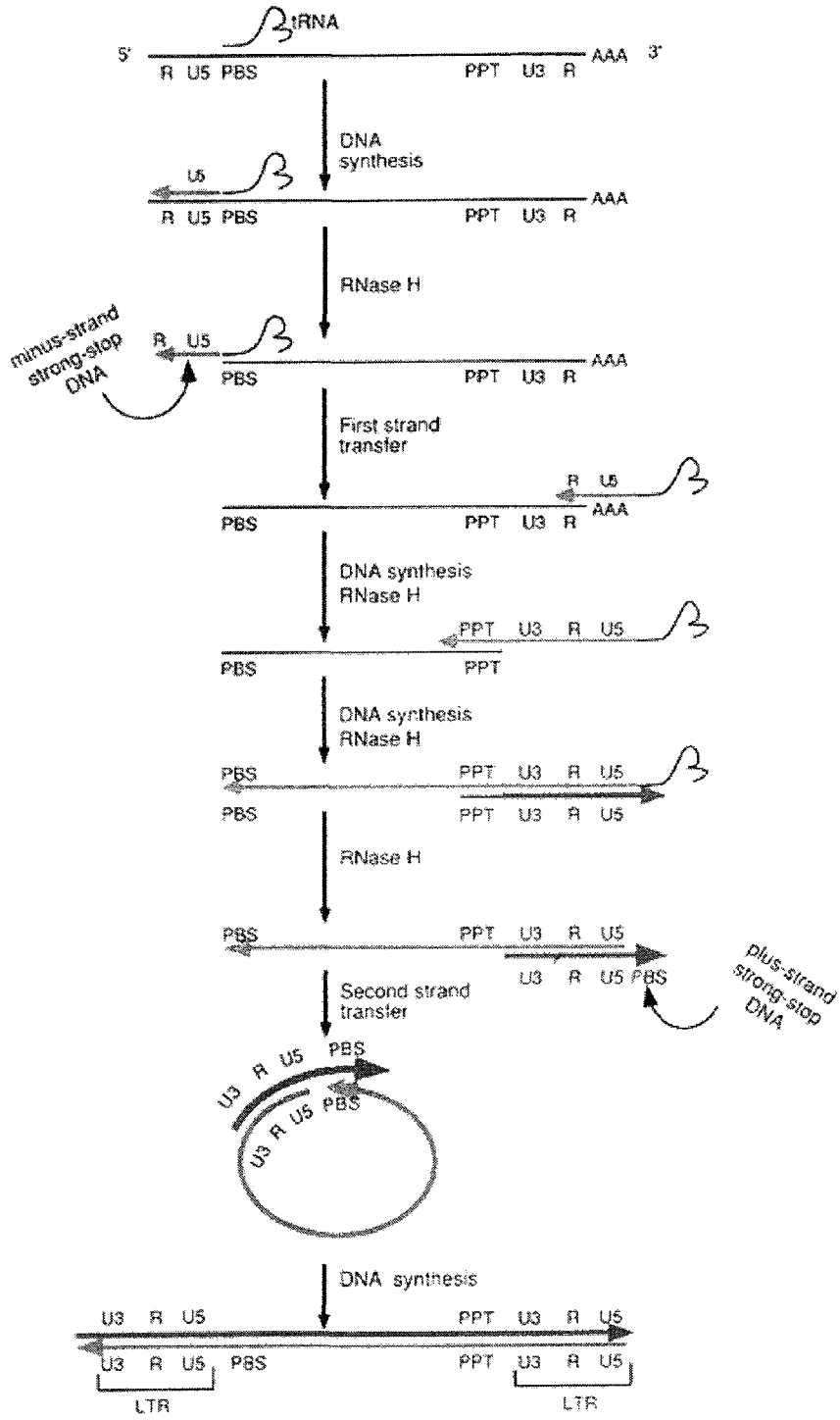


FIGURE 1.4. Mechanism of reverse transcription (adapted from (100)). Viral RNA is shown in Black. The minus strand DNA is in orange while the plus strand DNA is in red.

Genome Integration. The proviral genome is contained within the pre-integration complex (PIC) consisting of a number of viral proteins such as RT, IN, NC, MA and Vpr. Members of the lentiviridae family, including HIV-1, have evolved ways to import their genome into the host cell nucleus independently of nuclear membrane breakdown, allowing integration to occur in non-dividing cells. This process is largely mediated through Vpr and MA, which contain distinct nuclear localization sequences that enable PIC to be imported into the nucleus through the nuclear pores. The proviral DNA can then get integrated in non-specific sites of the host chromosome through the mediation of viral integrase. Viral integration sites can often be identified by the presence of short direct repeats flanking the viral genes (62,81,85,101).

Viral Gene expression. Once integrated, transcription and translation of the proviral DNA allow for viral gene expression and genome replication. Both viral and cellular proteins are essential for proper viral propagation (28,42). Cellular activation can generally trigger viral gene expression by recruiting transcription factors to the LTR region of the proviral DNA and allow for increased rates of transcription mediated by host RNA polymerase II (102). Other viral proteins such as Tat and Rev also contribute to the regulation of gene expression (28,42). The primary RNA transcript can either serve as viral genome and get packaged into the new virions; or serve as the transcript for viral protein translation (47). Levels of Rev protein present in the cell regulates the amount of RNA splicing where the full-length RNA transcript can remain unspliced, be singly spliced, or undergo multiple splicing events (2,103,104). The amount of splicing gives rise to the expression of different viral proteins, which in turn regulate the progress from early-phase to late-phase infection. For example, viral proteins Rev, Tat and Nef are translated from the doubly spliced RNA transcripts, which allows for their early expression. Vpr, Vif, Vpu, Gag, Pol and Env on the other hand, need Rev to transport their corresponding RNA

out of the nucleus and into the cytoplasm. As such, their expression must follow that of Rev and hence these proteins are referred to as late proteins (2,104) (Figure 1.5).

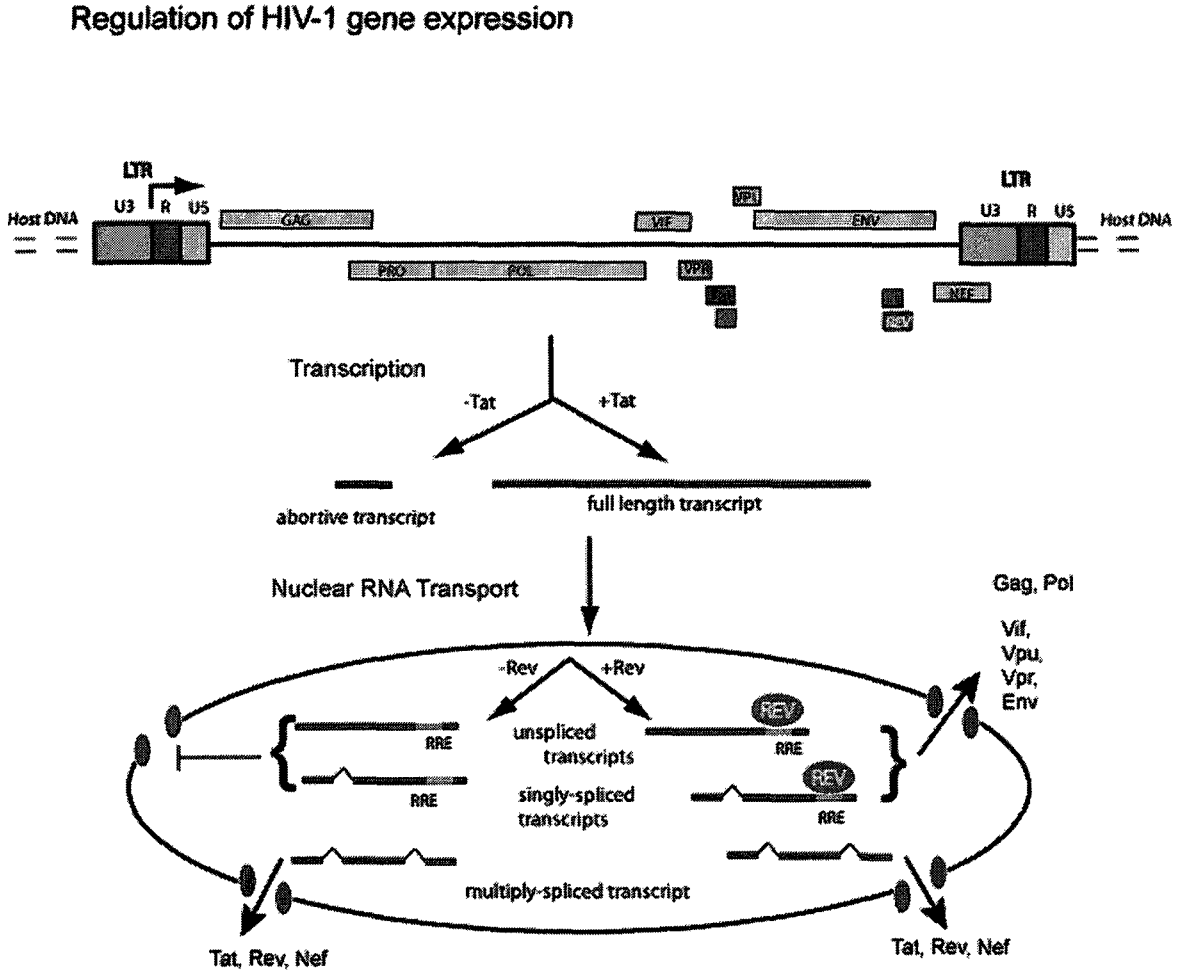


FIGURE 1.5. Regulation of gene expression during early and late HIV gene expression (adapted from (104)).

Viral Assembly and Maturation. Once all the viral proteins, and the RNA genome have been expressed, the viral packaging and budding can occur to give rise to progeny viral particles. HIV assembly occurs at the plasma membrane of the host cell where the Gag precursor polyprotein directs the assembly process (2). Assembly begins when the genomic RNA associates with Gag and Gag-pol polyproteins. Sequences in the NC region of Gag recognize the packaging sequence Ψ (psi) present in the full-length RNA genome and allow for the recruitment of the latter

(2,105,106) (Please refer to section 1.2.6. for a more detailed description of this process). The Gag and Gag-Pol precursor polyproteins are recruited into the assembling virion through interactions with various domains of Gag and Gag-Pol namely the C terminal region of CA (107) and regions in the IN and RT (64,108). For Gag and Gag-Pol membrane targeting to occur during virion assembly, a myristate moiety is added to the N-terminal of the MA and serves as an anchor for binding to the cellular membrane (32,33). It appears that the interaction domain of NC may further facilitate this process (109). The Env polyprotein moves through the ER and is transported to the cell membrane using the cell transport system. Budding of the viral particle begins with the association of the genomic RNA/ polyprotein complex with the Env glycoproteins inserted into the cell membrane (2,110).

The viral particle is subsequently released from the host cell and the viral maturation process is initiated. The virion particle initially released from the host cell remains immature until the packaged polyproteins molecules are cleaved to give rise to functional proteins. The protease enzyme (PR), as part of the Gag-Pol precursor polyproteins, first enables its activity through an auto-cleavage process. The PR enzyme then proceeds to cleave the rest of the polyproteins, giving rise to fully-formed viral proteins as well as core condensation. The step-wise process of PR-mediated cleavage eventually leads to the maturation of the virion into its infectious form (2,111,112).

1.2. NC protein in HIV-1: overview and structure

Retroviral NC proteins are nucleic acid-binding proteins with potent RNA-chaperoning properties, enabling important structural rearrangements that are required for genomic RNA replication and packaging, during virion assembly (113). NC has a high affinity for single stranded regions within RNA molecules (114) and strongly binds to sequences rich in G and U

(115). Retroviral NC was first isolated as part of ribonucleoprotein (RNP) complexes of Mulony Leukemoa virus (116). It was then shown that this stable RNP also contains viral dimeric genomic RNA, reverse transcriptase and cellular tRNA molecules, and that it can support DNA synthesis *in vitro* (117). Later on, Meric et al., found that NC is the major contributor in genomic RNA selection, packaging and dimerization in ASLV (118). A role of NC in genomic RNA dimerization of MLV was also observed by Gorelick et al.(119). These findings led to the conclusion that NC has nucleic-acid chaperoning properties that were later shown to be important for retroviral DNA synthesis (120). Retroviral NC proteins are small (usually less than 100 residues) and highly basic proteins generated after Pr55gag processing by viral protease, and except in the case of the spumavirus, they contain one (gammaretroviruses) or two zinc finger motifs with the highly conserved sequence CysX2CysX4HisX4Cys where X can be any amino acid (121,122) flanked by multiple basic residues (30). Each zinc finger coordinates one zinc ion through the CCHC-Zn interaction (123).

Mutation of zinc ion coordinating residues, like Cys23 reveals their critical role in maintaining the correct conformation of the protein (124). Zinc fingers from different retroviruses, in spite of sequence differences in the amino acid residues, show similar structures (125,126). In all retroviruses, not only the spacing of zinc binding residues is conserved, but the context residues in each finger are also highly conserved in each retroviral species (127,128)(Figure 1.6.).

Properties of retroviral NC proteins

Virus	(protein)	Size	N-terminal zinc finger	C-terminal zinc finger	Basic Residues	Acid Residues
RSV	(p12)	89 a.a.	CYTCGSPGHYNANC	CNLCNGMGHNAKNC	16 (18%)	3 (3.4%)
MLV	(p10)	60 a.a.	CAYCKEKGHWAKDC	N.A.	14 (23%)	8 (13%)
HTLV-1	(p15)	85 a.a.	CFRCGKAGHWSRDC	CPLCQDPTHWKRDC	11 (13%)	11 (13%)
HIV-1	(p7)	55 a.a.	CFNCGKEGHIAKNC	CWKCGKEGHQMKDC	15 (27%)	4 (7.3%)
SIV	(p8)	52 a.a.	CWNCGKEGHSARWC	CWKCGQMGHVMAKC	12 (23%)	1 (1.9%)
FIV	(p10)	66 a.a.	CFNCKKPGHLARQC	CNKCGKPGHVAAKC	14 (21%)	1 (1.5%)
PFV	(p75)	142 a.a.	N.A.	N.A.	27 (19%)	8 (5.6%)
Copia		43 a.a.	CHHCGREGHIKKDC	N.A.	14 (32%)	2 (4.6%)
Ty3		58 a.a.	CFYCKKEGHRLNEC	N.A.	18 (31%)	4 (6.9%)

Figure 1.6. Properties of retroviral NC proteins (adapted from (127)).

The HIV-1 NC protein consists of 55 amino acids, 15 of which are highly basic with only four acidic residues (figure 1.6). Residues 14 to 28 form zinc finger 1 while residues 36 to 49 form zinc finger 2 (30). Each of the zinc fingers contain an aromatic residue and the two motifs are linked by the basic sequence RAPRKKG present in the linker region (residues 29 to 35) (30) (Figure 1.8.A). The conformation and flexibility of the linker residues are responsible for the close proximity of the two fingers (129). Zinc fingers of HIV-1 NC are not functionally equivalent and substitution of either of the zinc fingers by its counterpart will disrupt their function (130). As seen by NMR studies, free NC has an ordered/ disordered structure: The N and C terminal regions of NC do not show a conserved structure, while both fingers form an ordered and stable structure when associated with a zinc ion. The flexible linker region allows for the globular conformation of the protein which brings residues Phe16 in the first zinc finger, and Trp37 in the second zinc finger, in close proximity to each other to form an hydrophobic

plateau (131) (Figure 1.8.B). Other hydrophobic residues, Ala25 from finger one and Met46 from finger 2 also contribute to the formation of this hydrophobic plateau (113). One of the first consequences of this hydrophobic plateau formation is the redistribution of the basic residues in the context of the fingers which is trivial for specific interaction with nucleic acid (132). Consequently, In HIV-1 NC, Lys 14, 20 and 38 on one hand, and Lys26, Arg29 and Arg 32, on the other hand, are clustered in the context of the Phe16 and Try37 aromatic residues (113). The formation of a transient globular hydrophobic structure was shown by observing weak and strong nuclear overhauser (NOE) effect between two fingers and was later confirmed by FRET and NMR studies (133). The presence of one proline residue in the linker region may also facilitate interaction between the two fingers (30).

NMR structures of NC attached to dACGCC or to an RNA resembling SL2 and SL3 hairpins from viral 5' un-translated region have already been obtained (134). According to the structural data obtained, the mode of interaction of NC with single-stranded DNA or RNA is roughly the same with respect to the amino acids involved and the globular structure of the zinc finger domain (135). After binding to an RNA oligonucleotide, the N-terminus of NC transforms into a 3 10 helix that lies in the major groove of the RNA (136). In the binding of NC to an RNA model of the SL3 stem-loop, F1 and F2 interact with G9 and G7, respectively, and the guanosine base packs within the hydrophobic plateau and the proximity of the fingers is reinforced (126). The NMR structures also show that the hydrophobic plateau at the surface of the zinc fingers represents the oligonucleotide-binding motif (113). The binding of NC triggers structural changes in the loop of SL2 and SL3, with no significant effect on the structure of the stem (126). Mutations that affect the globular structure of the hydrophobic plateau but not zinc ion binding residues also affect the function of the protein as shown in *in vitro* experiments (137).

Like other nucleic acid binding proteins, NC protein from all studied retroviruses have strong RNA and DNA chaperoning properties, which make them an indispensable component in almost all steps of the retroviral life cycle (138). RNA chaperone proteins can be found in all living organisms and display various functions in different steps of cellular life cycle like: maintenance of chromosome ends during cellular replication, DNA transcript, modification and splicing of preRNA in the nucleus, and mRNA translation and degradation (137). RNA molecules can be trapped in non-functional misfolded secondary and tertiary structures which may not be compatible with their proper role. RNA chaperone proteins help nucleic acid molecules reach their most thermodynamically stable conformation by resolving misfoldings and assisting with their aggregation. From this point of view, retroviral NC protein can be seen as a good representative of multifunctional RNA chaperones that drive the required genomic RNA structural rearrangements during the early and late phases of virus replication (113).

The chaperoning property of NC is partly related to its ability to destabilize secondary and tertiary structures of the nucleic acids (139). They also can direct annealing of complementary nucleic acid sequences. *In vitro* experiments have revealed that the NC protein needs both zinc fingers and basic residues to function properly. It is shown that basic residues in NC are mainly involved in nucleic acid aggregation processes while zinc fingers are more critical for effects on helix destabilization. These properties are also related to the proportion of NC/RNA as shown in *in vitro* experiments. According to this model, NC in limited concentrations, attaches to the RNA and forms a stable Ribonucleoprotein (RNP) complex. But when the proportion of NC in the complex goes high, it can carry out its chaperoning effect by destabilizing the preformed secondary and tertiary structures. It also helps the complementary sequences to anneal to their most thermodynamically stable form. On the other hand, by increasing the ratio of the NC

protein, it will unwind all secondary and tertiary structure of the RNA(137). Whether these reactions happen in the HIV-1 virus or not need to be elucidated.

NC is initially expressed as part of the Gag and Gag-Pol polyproteins (Figure 1.7). *In vitro* protease assays using recombinant HIV-1 protease and Pr55gag have demonstrated that in stage 1, Pr55gag is processed into MA–CA-p2 and NCp15. In stage 2, MA–CA-p2 and NCp15 are simultaneously cleaved into MA, CA-p2, NCp9 (NCp7–p1) and p6, a reaction that is roughly 10-fold slower than stage 1. Stage 3, which is roughly 35-fold slower than stage 2, leads to the appearance of CA and NCp7 at about the same time (50,140,141). Since it is very unlikely to have complete Pr55gag processing in the virus (142), and since the processing takes some time, it is likely that these intermediate proteins may play a role in the viral life cycle. NCp7 represents the dominant form of the protein found in mature viral particles (143,144). Overall, introducing a defect in the NC protein often resulted in modifications to viral structure and infectivity (121).

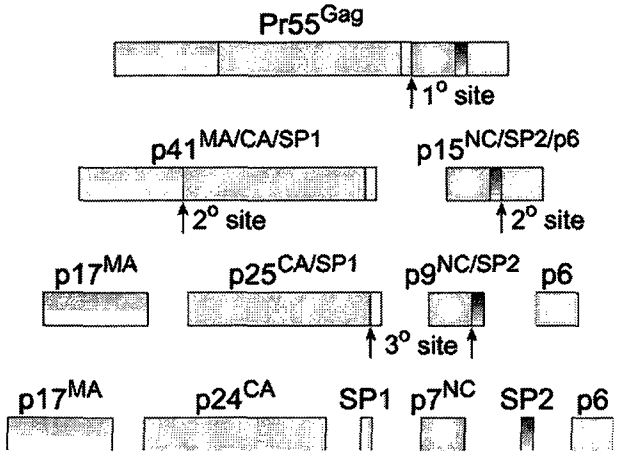
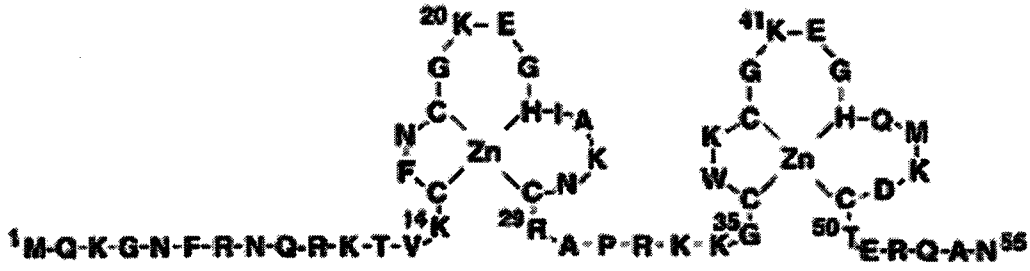


FIGURE 1.7. Stepwise proteolytic cleavage of HIV-1 NC (adapted from (127)).

A



B

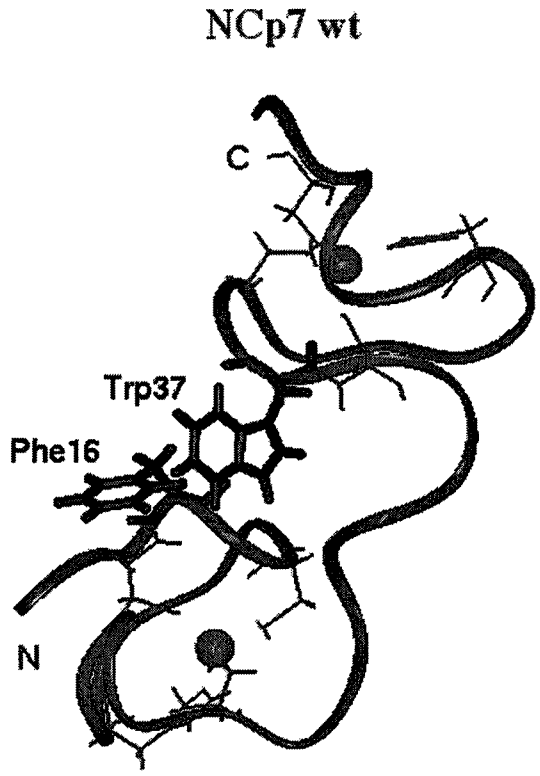


FIGURE 1.8. Sequence and Structure of HIV-1 NC protein (adapted from (145) and (121)).

In the following sections I will describe the role of NC in HIV-1 viral life cycle in two parts consisting of early and late phases of viral life cycle. HIV-1 viral genomic RNA dimerization

and the role of NC protein in genomic viral RNA dimerization will be discussed separately in the last section.

1.3. Role of NC in early viral life cycle

The early phase of HIV-1 viral life cycle begins with the attachment of an infectious virus to the cellular receptors CD4, CXCR4 and/or CCR5, which causes fusion of the viral membrane with the target cell membrane, allowing for viral core entry into the cytoplasm. The viral core then disassembles and reverse transcription transforms the single-stranded viral genomic RNA into a double-stranded viral DNA (vDNA). The proviral DNA then translocates to the nucleus and is integrated into the host genome, thereby completing the early phase of HIV infection (2).
In early stages of the HIV-1 life cycle, NC acts mostly in RT-mediated vDNA production and its integration by retroviral IN (integrase) inside the host genome. In all *in vitro* reactions mimicking early viral life cycle, NC appears as a nucleic acid chaperoning protein that helps nucleic acids reach their most thermodynamically stable arrangement. It is worth reminding that the same property is responsible for NC's function during a later phase of HIV-1 viral life cycle during which NC acts as a domain of Pr55gag polyprotein to promote gRNA packaging and viral assembly (146). The nucleic acid chaperoning properties of HIV-1 mature NC proteins in early viral life cycle have been extensively investigated using a wide variety of cell-free assay systems, as well as *in vitro* experiments (127).

In the following sections, first, I will briefly describe the role of NC in reverse transcriptase complex (RTC), with special attention to its interaction with RT. Its effect on HIV-1 genetic variability will be discussed separately and finally I will include a short review on the effect of

NC on transforming RTC to PIC and its role in integration of newly synthesized vDNA in host genome.

1.3.1. Role of NC in RTC

As previously mentioned, entry of the core into the cytoplasm of the cell results in core disassembly and RT-mediated transformation of the single stranded positive sense genomic RNA into double stranded vDNA. The initiation of reverse transcription in the cell is dependent on the intracellular availability of dNTPs (147) but it is also possible that the structure and shape of the free virus may interfere with efficient DNA production. Timely uncoating of the core looks important for viral infectivity as seen by low infectivity of HIV-1 mutants that have more stable viral cores (148). Dr. Gorelick and colleagues have recently reported that accelerated core disassembly and premature RT initiation may be responsible for vDNA instability and consequently low infectivity of viral mutants containing point mutations in CCHC Zn containing residues (127).

The RTC (containing dimerized RNA genome, $tRNA^{Lys3}$ primer, MA, p7NC, RT, IN, PR and Vpr) is a complex that initiates reverse transcription and produces the double stranded vDNA. The pre-integration complex (PIC), which contains dsvDNA, MA, NC, IN and Vpr, is responsible for the integration of full-length vDNA into the cell genome. RTC and PIC have dynamic characteristics, this property makes their isolation from the cells and differentiation from each other challenging and ambigious. But the vDNA detected 8 hours post-infection in the nucleus of the cell is generally considered to be PIC vDNA (127). Precise details of the interactions behind this transformation remain to be elucidated. In the following sections, I will review the role of NC protein in the major steps of the transformation of RTC into PIC.

1.3.2. Role of NC during Reverse Transcription

Genomic RNA reverse transcription can be initiated as soon as the viral particle is formed by host cell $tRNA^{Lys3}$ annealing to the PBS region of viral gRNA and initiation of the cDNA synthesis (149,147). After budding, a small minority of the viruses (1 in 1000) have minus-strand strong stop DNA, but their significance in the viral life cycle is not clear (150). When the released virion infects a new cell, the RNP core structure is released into the cytoplasm and cDNA synthesis can resume (Figure 1.3). The two obligatory strand transfers during reverse transcription and viral genomic RNA secondary structures are the major obstacles during reverse transcription. It has been shown in both *in vitro* and *in vivo* experiments that NC increases the efficiency of all steps of reverse transcription, from tRNA placement to vDNA protection (127). The role of NC in this process can be reviewed as the following steps:

Initiation: Observations from *in vitro* studies show that NC can chaperone the $tRNA^{Lys3}$ molecule towards PBS by partially disrupting the primary structure of the $tRNA^{Lys3}$ and annealing its 3' 18 nucleotides to the corresponding complementary sequences in PBS (139,147) (Figure 1.9.A.). During this step and based on *in vitro* studies, NC can greatly enhance the interaction of the two molecules which is also dependent on the simultaneous intecractions of NC protein and both genomic RNA and primer $tRNA^{Lys3}$(61,151,228). In the absence of the NC protein, heat can be used to facilitate the same interactions indicating the importance of the NC protein to carry out this step at physiologic temperature (152). Molecular interactions between RT and NC from one hand and affinity of RT p66-p51 for primer $tRNA^{Lys3}$ seems to be important for the initiation of reverse transcription (151,147). The incorporation of RT and Primer $tRNA^{Lys\,3}$ are explained in the viral assembly section.

Minus Strand Strong Stop DNA: The newly synthesized DNA from the tRNA primer to the 5' end of the genomic RNA is minus-strand strong stop DNA (Figure 1.4). As DNA gets transcribed, RNase H activity of p66 subunit from RT heterodimer cleaves the RNA portion of the heteroduplex to small fragments, leaving a single stranded cDNA that is complementary to the other R region located at the 3' region of genomic RNA, but this region is also capable of self-priming (153). It is worthnoting that , in *in vitro*, NC enhances the RNase H-mediated RNA degradation reaction mediaterd by HIV-1 RT molecules, (154). This region (R-U5 sequences also called (cTAR)) contains multiple secondary structures that may interfere with proper reverse transcription. In this region, the TAR hairpin has a particularly stable structure that also plays a role during viral RNA transcription from proviral DNA by recruiting Tat accessory protein to the transcriptional machinery (155). Through its chaperoning effect, NC can destabilize these secondary structures and improve reverse transcription processivity (156). The helix unwinding activity of NC inhibits self-priming of the minus stand strong stop DNA while promoting its annealing with the 3' end R RNA of one or the other of the 2 genomic RNAs. This annealing is called first obligatory strand transfer (157).

First cDNA transfer: The first strand transfer event during reverse transcription of viral DNA needs both the newly synthesized cTAR and viral 3' end TAR sequences (158). NC destabilizes cTAR and facilitates the opening of its terminal base pairs (158). Destabilization of the TAR sequence at 3' end of viral genomic RNA by NC is not as efficient as cTAR opening in *in vitro* experiments probably due to its higher stability (159). There is a strict requirement for both zinc fingers of NC protein in these destabilizing experiments which seems to implicate the interaction of cTAR and TAR with the NC hydrophobic plateau (129). Basic residues of the NC are mostly involved in hybridization of cTAR and TAR (158). This transfer can be intramolecular or

intermolecular. In the later case, both RNA molecules should be in the same ribonucleoprotein complex which can be provided by dimeric genomic RNA of HIV-1. Intermolecular strand transfer events increase the possibility of recombination during cDNA synthesis. After the first obligatory strand transfer, RT resumes (-) strand DNA transcription and at the same time the RNase H activity of RT continues to degrade the RNA template except for a small region called the polypurine track. This small region will be used to initiate plus strand DNA synthesis.

Plus-strand DNA Strand Transfer: During the final obligatory strand transfer during cDNA synthesis, minus strand PBS region hybridizes to its plus-strand counterpart. This step, like minus strand transfer, can happen both intra or intermolecularly. Intra-strand transfer forms a circular intermediate while inter-strand transfer results in an end-to end linear product (160). Even though the secondary structures of the short (18 nt) nucleic acids involved in this transfer are not complex compared to the first transfer, NC still enhances this strand transfer reaction (157) (Figure 1.9.B.). Interestingly the zinc finger structure of the NC can be eliminated with little effect on the plus-strand DNA transfer. But a NC mutant with both zinc fingers containing SSHS mutation is severely compromised in minus-strand transfer step in *in vitro* transcription system (157). It is worth mentioning that additional strand transfers, involving other than the two obligatory sites, can happen during reverse transcription. These strand transfers are imposed during reverse transcription of positive or negative strand DNAs either by genomic RNA secondary structure or nicks in the genomic RNA. NC has been shown to facilitate these strand-transfers which contribute to boosting the rate of recombination observed between the two RNA genomes (113,146). During the final step of reverse transcription, minus strand DNA synthesis is extended by using the newly formed plus strand as a template, leading to the duplication of the 3'LTR at the cDNA 5'end. RT by itself is capable of carrying out this process, but as in the

previous steps in cDNA synthesis, this step is also enhanced by the presence of NC (161). Although several groups have recently focused on understanding the role of NC in RTC both *in vitro* and cell culture experiments, little is understood about the mechanism of transformation of RTC to PIC. By using an *in vitro* model of RTC, Bampi et al., recently showed that NC can facilitate nucleotide excision repair by reverse transcriptase (162). It is expected that the coming years will provide us with more information regarding the role of cellular proteins in RTC and PIC formation and function.

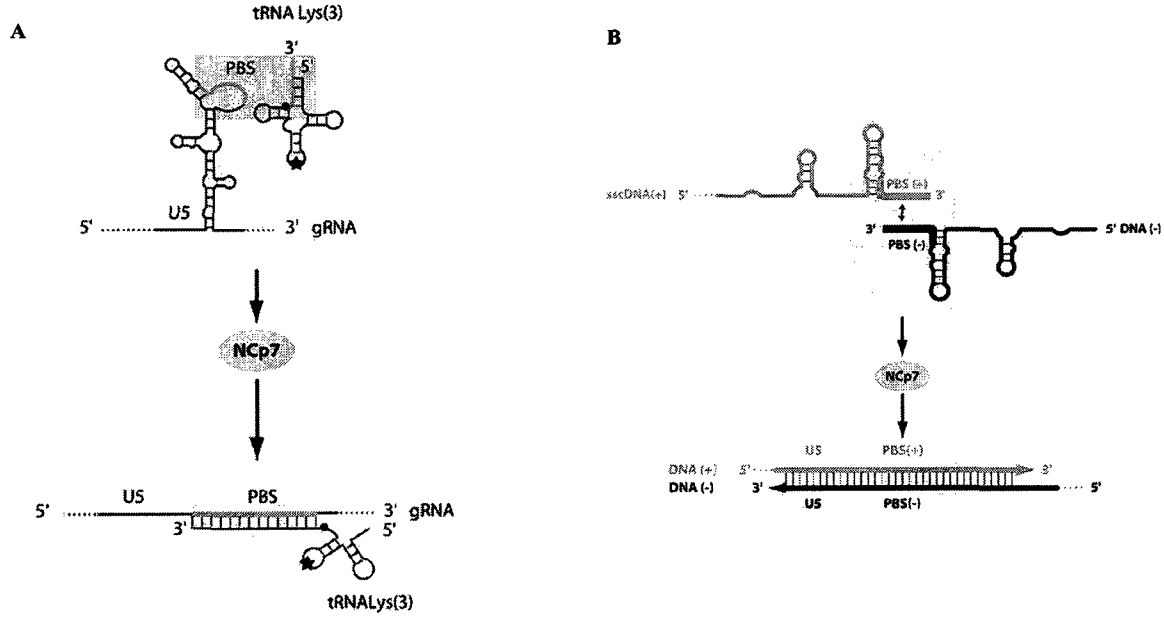


FIGURE 1.9. Role of HIV-1 NC during reverse transcription. **A.** NC mediates the binding of $tRNA^{Lys3}$ to PBS. **B.** NC chaperons the hybridization of PBS regions during strand transfer events (adapted from (147).

1.3.3. Role of NC in Genetic Variability

HIV consists of a group of highly diverse quasispecies (2). Two major contributors to the high genetic variability of HIV are the high error rate of reverse transcription and its desoxynucleotidyl transferase activity (Figure 1.10). RT can start vDNA synthesis at false sites and misincorporate nucleotides in the vDNA. But because it lacks proofreading activity, it

cannot correct these mistakes (2,163). HIV-1 RT enzyme, like other retroviral reverse transcriptases, has desoxynucleotidyl transferase activity, where one to three nucleotides are added to the end of the newly synthesized DNA strand in a template-independent manner (164). This can happen during every strand transfer event that happens in each round of vDNA synthesis. The transferred strand containing the extra nucleotides increases the amount of mismatch extension and hence increases genetic diversity (165). While this reaction appears to be slow *in vitro*, the presence of NC and its interactions with RT appear to significantly enhance mismatch extension (166) (Figure 1.10.B). During reverse transcription, RT can dissociate from one synthesized strand and continue synthesis on the RNA strand of the second genome giving rise to recombinant proviral DNA (167). This process appears to be enhanced through two simultaneous, yet distinct mechanisms. In the first process, RT-mediated pausing during DNA synthesis increases RNase H activity which in turn increases the transfer of the ssDNA to a new RNA template (168). The second mechanism occurs when an incoming RNA template actively displaces the donor template, forcing DNA synthesis to continue on the former. NC appears to actively support the latter mechanism and hence increase the number of recombination events (168) (Figure 1.10.A.).

Due to the high error rate during reverse transcription, several copies of the HIV genome can simultaneously exist in an infected cell. Furthermore, the cell can be co-infected with genetically distinct HIV particles, as different viral proviruses have been identified in the lymph node of patients during the acute phase of the disease. Together, these factors contribute to increased genetic recombination during viral replication (2,147).

Because NC plays a role during reverse transcription, it can contribute to increased viral genetic diversity through different mechanisms (147). Based on the observation that NC can enhance

RT-mediated excision repair activity (162), we observe that NC has a dual role during viral DNA synthesis. It can ensure faithful vDNA synthesis which contributes to viral infectivity and viability through enhancing excision. But on the other hand, it can promote recombination which enables the virus to have sufficient genetic variability in order to escape immune response and produce resistance strains during HAART treatment.

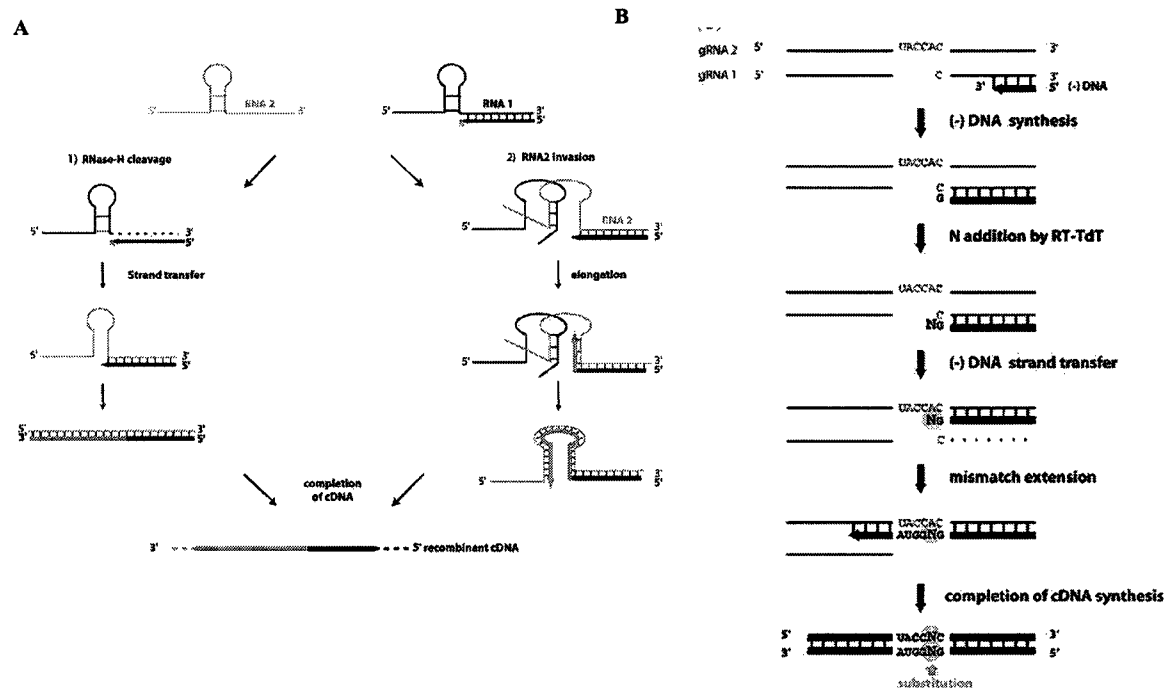


FIGURE 1.10. Role of HIV-1 NC in promoting genetic diversity. A. NC enhances the active displacement of RNA templates resulting in template switching during reverse transcription. B. NC promotes the terminal transferase activity of RT, resulting in increased mismatch extension (figure adapted from (147)).

1.3.4. Role of NC in transforming RTC to PIC:

After virus entry and viral core uncoating, genomic RNA, as well as cDNA in both RTC and PIC complexes, are likely to be targeted by cellular nucleases. *In vitro* experiments have shown that NC can bind to and protect nucleic acids against nucleases (156,169). Reduced attachment of NC mutant proteins to nucleic acids in the same experimental situation may explain their low

protective effect against nucleases. On the other hand, HIV-1 proviral vectors with NC mutations in cell culture experiments, yield cDNAs with unstable ends that are not capable of integration (170-172). All these experiments indicate that NC may play a role in protection of viral gRNA and vDNA inside the cell. The concentration and kinetics of attachment of the NC to RTC is dynamic and changes as the viral nucleic acid metamorphoses from ssRNA in RTC to dsDNA in PIC. Because attachment of NC to ssDNA is different from its attachment to dsDNA, this leads to the accumulation of more NC on the central flap of the vDNA and stabilizes integrase molecules at both ends of the vDNA (173). This model was strengthened by *in vitro* experiments showing that NC attaches to ssDNA more efficiently than to dsDNA (174). In addition, TEM visualization of dsDNA containing a central flap demonstrated NC accumulation on this region (175). This rearrangement of IN and NC protein may help in nuclear localization of the vDNA and increase the efficiency of integration. Even though in mature 48 hours old HIV-1 viruses, 95% of the NC is in the form of NCp7 (30), it is likely that other intermediate NC containing Pr55gag processing proteins (p15, p9) may play a role in these processes. For instance, *in vitro* experiments has shown that NCp9 is more potent that NCp7 in aggregation studies and carry out *in vitro* integration interactions better than NCp7 (176).

1.3.5. Role of NC in vDNA integration

The newly synthesized double stranded proviral DNA in the pre-integration complex though formed in the cytoplasm, can be detected in the nucleus of infected cells 8 hours post infection (127). NC, along with the integrase enzyme, enhances proviral integration into the host chromosome. The NC molecule thoroughly coats the proviral DNA (177), protecting the nucleic acid from nuclease-mediated degradation which is crucial for proper integration (172). In addition, NC coating appears to enhance end-to-end interactions between DNA molecules. This

leads to blunt end ligation reaction during integration (178). In addition to coating the DNA molecule, NC appears to increase the recruitment of IN to the PIC, hence enhancing DNA-IN interactions (147). NC mutants lacking the N and C terminal region were not able to form the ternary complex while when the zinc fingers were knocked out the NC-DNA complex was able to form, although IN recruitment was significantly reduced. Hence, it appears that NC mediates the interaction between IN and DNA (147).

In conclusion, during the early phase of the viral life cycle, NC protein acts mainly as a nucleic acid chaperoning protein by facilitating the multi step enzymatic interactions of reverse transcription and integration as shown by *in vivo* and *in vitro* experiments. NC also plays a key role in gRNA and vDNA protection during these processes. The precise role of NC in RTC and PIC, especially any role for cellular elements in theses interactions, still remains to be elucidated.

1.4. Role of NC in late viral life cycle

The late phase of the HIV-1 life cycle starts with synthesis of viral spliced and unspliced RNAs and their translation using cellular translational machinery. The HIV-1 provirus inserted in the host cell genome produces three classes of viral RNAs: multiply spliced, singly spliced and unspliced viral RNAs (179). Multiply spliced viral RNAs can exit the nucleus and their translation by cellular ribosome yield accessory proteins Tat, Rev and Nef. After entering the nucleus, Tat recruits CDK9 through Cycline D, and enhances viral RNA Production(180). The Rev protein attaches to the Rev Response Element, located at the 3' end of the unspliced and singly spliced RNAs, and using cellular Exportin 1, it brings the unspliced and singly spliced RNAs inside the cytoplasm (179,181). Singly spliced RNAs are translated in RER and, after multiple post-translational modifications and processing by cellular protease. The resulting viral

proteins (SU and TM glycoproteins) are translocated to the cell membrane by the Golgi complex for further incorporation in the budding virus during assembly process (182).

Unspliced RNAs, on the other hand, are translated by free cytoplasmic ribosomes to produce the two major structural proteins: Pr55gag and Pr160gag-pol. Pr55gag is the major structural protein and is sufficient for producing viral like particle (VLPs) by transfeced cells. The NC protein domain of Pr55gag, by attaching to the Psi-elements (ψ) in the 5'untranslated region of viral RNA, changes its fate from a mRNA to gRNA which is capable of encapsidation in the assembling virions (120). Pr55gag complexed with RNA can incorporate Pr55gag-pol polyproteins (124). With or without viral RNA, it is also transferred to the cell membrane where viral assembly takes place in T cells. The assembly site is cell specific: it happens in the multi-vesicular bodies monocytes (183). Viral particles bud from specific microdomains of cellular membrane named lipid rafts which are rich in sphingomylein and cholesterol. They also contain an envelope after budding which is from the cellular membrane. Immature newly released viral particles are composed of ~5000 molecules of Pr55gag and appear spherical in the electron microscope, with no conical cores (184). Activation of viral protease in the released virus cleaves structural proteins and makes the virus mature-looking, and capable of initiating a new round of infection (30,142,153).

In the following section, I will review our recent understanding of the role of NC protein in the late phase of viral life cycle.

1.4.1. Pr55gag synthesis and multimerization

Pr55gag is translated from full length and unspliced viral RNA, using cellular translational machinery. Its translation is believed to take place by two different mechanisms: 1) viral RNA acts as a 5' capped mRNA which gets translated by free cellular ribosomes as other cellular mRNAs. 2) Though internal ribosome entry signal or IRES that is in the 5'untranslated region of the viral RNA which is believed during cell cycle arrest in G2/M phase that are brought by Vpr (185-187). Pr55gag alone is enough to produce VLPs from the infected cells, to incorporate RNA and Pr160gagpol in the assembling complex and also to bring membrane proteins SU and TM in the budding particles (188,189). It can self-assemble into VLPs by ordered multi-merization of Pr55gag monomers to produce a spherical shell, which forms the structural framework of the immature virus particle (132). Multiple interactions between Pr55gag molecules are responsible for its multimerization.

MA in the N-terminus of Pr55gag by its basic residues and covalently attached myristic acid is the major determinant to drive and bind Pr55gag to the cell membrane. It has been proposed that MA trimerization during Gag multimerization can expose myristic acid to interact with the cellular membrane. The contribution of MA in viral assembly is under scrutiny after several researchers showed that its deletion does not totally inhibit VLPs formation (190,191). The I domain, mapped to the CTD of CA protein, p2 and N-terminus of NC protein promotes viral assembly by multiple functions including, specific and non-specific RNA attachment, Pr55gag multimerization and Pr160gag-pol incorporation (192,193). NTD region from CA does not seem to play major role in Gag multimerization, since its deletion does eliminate VLPs production (194). The ability of CA protein to from dimers is believed to be important for Gag multimerization. This region also contains highly conserved MHR sequences that are important

in viral assembly (195). During viral assembly, cellular cyclophilin A is incorporated in the virion through its interaction with praline-rich loop in the CA-NTD. Its precise action in the virus is not completely known, but recently it has been proposed that it has a role in viral early life cycle by protecting the virus from cellular restriction factor TRIM5α (196). The minimal amino acid residues necessary for I domain function were contributed to the basic residues in the N-terminus segment of NC protein (109,192). Hence, the major function of NC protein representing I domain in viral assembly, comes from its ability to bind RNA. RNA in this regard acts as structural scaffold that increases Gag concentration and facilitates their interaction and multimerization (192,197). Specific preference of NC protein for full-length viral RNA promotes gRNA packaging, but in the absence of viral full-length RNA, other non-specific nucleic acids can play the same role (192,197). The L domain in the C-terminal part of Pr55gag helps viral particles bud off the cell by recruiting Tsg101, which is a host component of the cellular endocytosis machinery with the help of PT/SAP motif in the N-terminal segment of p6 (27,48). Any mutations in this motif will cause a defect in the release of the virus, with characteristic attachment of the viruses to the cell membrane with a thin stalk (198).

In the following section, I will briefly review HIV-1 viral assembly with emphasis on the role of NC protein.

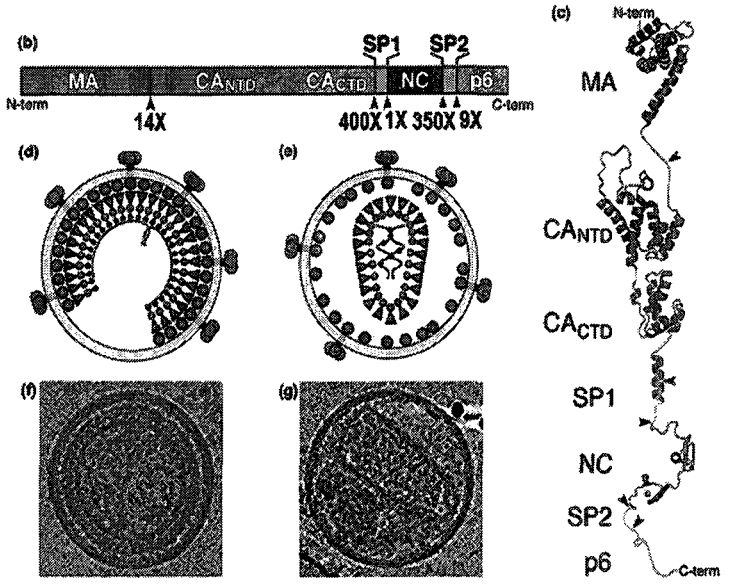


Figure. 1.11. Viral assembly, adapted from (87).

1.4.2. Role of NC in viral assembly

Pr55gag polyprotein alone can form extracellular virus-like particles in the absence of other viral proteins (189) and *in vitro* experiments have shown that they can spontaneously assemble into virus like particles (VLPs) (199). NC is a key element for HIV-1 assembly and the absence of NC results in no, or very little production of VLPs (109,200,201). However, in protease-negative mutants devoid of NC, VLPs are produced. This led to the proposal that NC is mostly important for viral stability rather than viral assembly (201). Because the rescue of virion production by inactivating protease enzyme has not been reproducible in other cell types, it is speculated that this phenomenon is a cell type-specific effect. During Gag translation, as Pr55gag accumulates, it is possible they can bind to specific sequences in viral RNAs through their NC domain (202,203). This can be supported by *in vitro* studies that show Pr55gag can bind tenfold stronger than NC to the RNA sequences mimicking viral stem loops (204). Increased binding of Pr55gag

proteins to stem loops SL1-SL4 in 5' untranslated region of the full length viral RNA can remodel its secondary and tertiary structure through the chaperoning property of NC protein embedded in the Pr55gag protein, transforming a mRNA to viral gRNA and reorient it to dimerization and packaging and starting virus assembly (47,105,205). In this model of virus assembly, the genomic RNA acts as a platform for the recruitment of Pr55gag molecules via specific interactions between NC and the SL1-SL4packaging/dimerization signals (113). According to this model, it is conceivable that binding of NC from Pr55gag to RNA, can bring Gag molecules into a concentrated environment and promote their interaction and multi-merization. This model is also supported by several observations showing that VLPs derived from Pr55gag molecules incorporate cellular RNAs in the absence of viral gRNA (206) (207). RNA/Pr55gag assembly brings other viral and cellular elements to the complex and helps them relocate to cellular membrane for complete viral assembly. In HIV-1, basic residues of MA also contribute to RNA binding and virion assembly (208,209).

1.4.3. Role of NC in genomic RNA packaging

NC plays a key role in specifically recognizing and packaging the genomic RNA into the virus particle (210,211). NC regulates both the efficiency and specificity of genome RNA packaging (132). This is mostly achieved through highly specific interactions between HIV-1 NC and the Ψ sequence in the 5' leader region of viral gRNA. It has been shown that mutations in either zing fingers of HIV-1 NC protein or its basic residues will decrease the amount of incorporated gRNA in the virions (30,192). After starting Pr55gag translation from full-length RNA, it is proposed that NC as a domain of newly synthesized polyprotein can recognize and change the fate of full-length RNA from an mRNA capable of translation, to a gRNA which becomes

incorporated into the virus as genomic material (Figure 1.11) (212,213). The segment of the RNA genome involved in packaging spans hundreds of nucleotides and includes the entire 5' untranslated region (5' UTR) and half of the *gag* gene (132,210). In this area, four stem loops (SL1-4), are independently important in viral gRNA packaging (214). Among these, SL2 and SL3 are believed to be the major determinants of genomic RNA packaging, as *in vitro* experiments have shown higher affinity of NC protein for these stem loops (121,145). Deletion or disruption of SL3 results in a large reduction of genome packaging. However the deletion of multiple stem loops result in even higher disruption of packaging suggesting overlapping roles the stem loops. For example, SL2 has also been extensively implicated in NC interactions and genome packaging where SL2 also binds NC with high affinity even though the specific molecular interactions may be different from that of SL3 (102,133). The crystal structures of complexes NC-SL2 and NC-SL3 are depicted in figure 1.12.

SL2 also contains major splice donor site that may have a role in specifically encapsidation of unspliced full-length RNA in the virus (210) and SL1 has been identified as DIS (221).

NMR studies, as well as mutational studies have identified that each SL region can bind a single NC molecule with strong affinity and that the NC molecule preferentially binds to single-stranded sequences (102,222). The GXG motif present in all of the four stem loops was identified as the major factor in NC binding (86, 196, 211).

These studies also have implicated the Trp37 residue in NC in providing stacking interactions with the guanines of the GXG motif (Figure 1.14). Ionic interactions between basic amino acid residues of the protein with the phosphate groups present in the stem loops further stabilize the binding of the two molecules (102). However, a decrease in flexibility of this interaction may result in reduced binding and may account for the low affinity of SL4 for NC (194).

Analysis of the NC protein attached to an RNA molecule mimicking SL2 and SL3 region of 5' end from HIV-1 untranslated region revealed that both zinc fingers are responsible for specific interactions with ψ-site sequences on the gRNA, but the basic residues interact more non-specifically by strengthening the protein-RNA interaction (106,136). Deleting the 5' end ψ-site of the viral RNA or mutations of the NC domain from Gag will disrupt specific packaging of HIV-1 gRNA, although packaging of cellular RNAs is supported by intact NC region of the Pr55gag in the former and by basic residues of MA, in the latter mutants (215). According the these finding, a model has been proposed in which the tertiary structure resulting from SL1 dimer formation and SL4 stabilization exposes SL3 and SL4 for high-affinity NC binding (132,216).

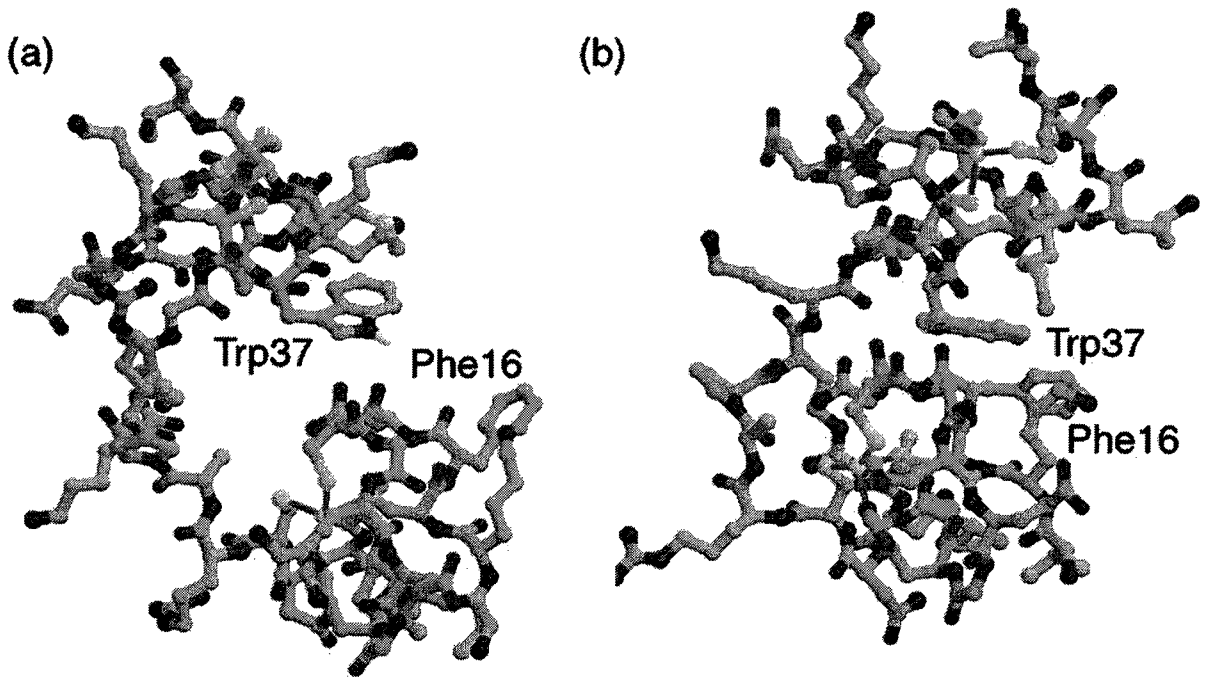


FIGURE 1.12. Crystal structure of HIV-1 NC with (a) SL2 and (b) SL3 (145).

1.4.4. HIV-1 viral assembly on the membrane

The primary site of HIV-1 viral assembly in CD4+ T lymphocytes is the plasma membrane (207). In macrophage derived monocytes, assembly and budding takes place at the late endosome/MVB (multi-vesicular bodies) (217). This implies that undiscovered host-cell factors may have a role in the site of viral assembly. As recently discovered, MA domain of Gag interacts with phosphoinositide phosphatidylinositol-(4,5)-bisphosphate [PIP(4,5)P2] of cell membrane and its alteration can redirect viral assembly from cytoplasmic membrane to multi vesicular bodies (218,219). The precise mechanism of this process is a hot topic in retrovirology research. In the following section, I will use the cell membrane assembly model to review viral assembly and viral budding. One of the first consequences of Gag multimerization is incorporation of other viral (e.g.Vpr, Vif) or cellular elements (e.g. tRNA lys, APOBEC3G, Tsg101) in the complex to form an infectious particle. Viral enzymes (PR, RT and IN) are translated in Pr160gag-pol polyprotein by -1 ribosomal frameshifting occurring in 5-10% of Gag synthesis (53). Gag-pol incorporates in the assembling particles through interactions with the CA domains of Gag and Gag-Pol (220). Interestingly, deleting Gag domains from Gag-Pol does not totally eliminate its incorporation in the assembling virions (221). The same authors also showed a role for p6 in Gag-Pol incorporation, as deleting of p6 only reduces its incorporation 4-5 times (221). RNA is also required for stable Gag/Gag-Pol complex formation (222), as mutations of NC domain in Gag that reduce its affinity to RNA, will also inhibit stable Gag-Pol formation. But the same mutations in Gag-Pol does not interfere with complex formation (222). Overall it appears that Gag-Pol does not directly interact with RNA but the formation of Gag/Gag-Pol complexes requires RNA-facilitated Gag multimerization (197,199). The proportion of Gag/Gag-Pol is important for infectious particle formation and genomic RNA dimerization (223).

Gag/Gag-Pol complexes are rapidly and almost completely associated with the host cell membrane while only 30% of newly synthesized Gag-Gag is targeted to the membrane (224). Selective incorporation of $tRNA^{Lys}$ into virions occurs during particle assembly. One of the $tRNA^{Lys}$ ($tRNA^{Lys3}$) will serve as the primer for the initiation of vDNA synthesis through reverse transcription (225). HIV-1 incorporates $tRNA^{Lys}$ into the assembling virion by co-assembly of cognate aminoacyl tRNA synthetase (226). Initial Gag/Gag-Pol/RNA complex formed after Gag multimerization might represent the $tRNA^{Lys}$ packaging complex (227). According to a model proposed by the same authors, this complex interacts with a $tRNA^{lys}$/ LysRS complex, with Gag interacting with LysRS, and Gag-Pol interacting with $tRNA^{lys}$ (227). 18 nucleotides of the incorporated $tRNA^{Lys3}$ molecule is complementary to the primer-binding site (PBS) on the HIV-1 viral genome. By using *in vitro* assays, several investigators have shown that HIV-1 NC, both in the precursor (Pr55Gag) and mature (p7NC) forms, is capable of annealing $tRNA^{Lys3}$ to its corresponding sequences on gRNA (134,203). Primer $tRNA^{Lys3}$ placement is also associated with the NC's nucleic acid chaperoning activity which is not interrupted by disrupted zinc fingers (228).

Proteins Env, Vif and Vpr are translated from single spliced viral RNA and are required for infectious viral formation. Vpr incorporation occurs through its interaction with p6 (229) and NC (172,205). In addition to its participation in the nuclear import of newly synthesized vDNA during early phase of viral life cycle, Vpr arrests the infected cell in G2/M phase of life cycle (230).

APOBEC3G, belongs to the cellular Cytidine deaminase-editing enzymes family and has anti-HIV activity (231). Viral Vif protein counteracts the anti-viral activity of APOBEC3G by two means: First, by directly interacting with APOBEC3G in the cytoplasm, promotes its proteasome

directed degradation and subsequently prohibits its incorporation in the assembling virus (208) and secondly in case of incorporation of APOBEC3G in the virus interfere with its deaminase activity (209). It is interesting to know that both APOBEC3G and Vif incorporation can be mediated by NC (232,233).Env glyproteins are synthesized by ribosomes of RER as a larger precursor (gp160), and undergoes extensive post-translational modifications and become glycozylated, form trimers and after being cleaved by cellular enzymes yield gp120 and gp 41 Env subunits (207). These subunits move to the cellular membrane through Golgi complex of the cell. Their distribution in the cell membrane is not random as their concentration is higher in the specific sub-domains of the cytoplasmic membrane, namely in lipids rafts (234). Rafts resistant to Triton-X100 have been best characterized and are described as a platform for assembly and budding of a variety of envelope viruses such as measles, influenza and HIV-1 (224). During viral assembly, interactions between MA from Gag, and cytoplasmic tail of gp41 promotes the incorporation Gag in the assembling particle (235). There is cell type specificity in the interaction of gp41 with MA, since Env protein truncation does not interfere totally with its incorporation in the assembling virus implying the involvement of a possible cell factor element in this process. Assembled immature virus buds off the cell, hijacking cellular machinery that is normally used to create vesicles that bud into the late-endosomal MVBs, as explained in the previous section (132).

1.4.5. Role of NC in viral maturation

The virus initially assembles and buds from cells as a noninfectious, immature, spherical particle that is organized by a layer of Gag proteins that are associated with the inner viral membrane

(Figure 1.11) (87). Structural studies of immature virus particles shows MA domain of the Gag polyprotein is attached to the cell derived viral membrane and the C-terminus of the Gag is radiated to the center. This arrangement gives a doughnut shape to immature virus. Protease-mediated processing of the Gag and Gag-pol polyproteins, changes this arrangement and a conical core structure assembles inside of the viral particle (236,237). Gag and Gag-Pol processing is achieved by the viral protease that is originally part of Pr160gag–pol and is active in both this precursor form and as a processed product (130,212,213). Even though the exact timing of this processing remains uncertain, studies with virions 10 second after release shows almost complete protein processing (108). Premature activation of PR or its over-expression has negative effect on viral assembly and leads to non-infectious viral particle formation (214). Protease inactivation also reduces viral production in cell culture experiments (108). PR activity in Gag-Pol is responsible for its release from the precursor protein. PR is active as dimer and both monomers should be intact to have an active enzyme (30). The first cleavage site in Gag-Pol, like Gag, is intramolucalar and occurs between p2 and NC to release MA-CA-p2 in the N-terminus, and NC-TF-PR-RT-IN in the C-terminal (46,212). Subsequent processing of theses intermediate proteins release individual structural proteins and cause viral rearrangement and restores viral infectivity. NC mutant HIV-1 particles show defect in the Gag processing which has been shown by western blot studies of the purified viral particles (30). For instance, mutation S3E which replace three basic residues of the linker in HIV-1 NC protein abolishes core formation inside the virions (238), suggesting that interaction between NC and RNA influences virion morphology. Other NC mutations generate HIV-1 viruses which are 80% immature in morphology, vs 3% immature in WT (192,239). Mutation in 5'UTR region of the viral genomic

RNA can also inhibit complete viral gag protein processing. These viral particles also lack mature conical core when visualized with Electron microscopy (240).

Viral maturation can be studied in three different ways: 1) detection intermediate or end products of Gag and Gag-Pol processing by western blotting, 2) Study of the conical core formation, as release of free CA is pre-requisite to the conical core formation and finally 3) by studying viral genomic RNA dimerization as an indicator of viral maturation. In the following section I will discuss HIV-1 viral genomic RNA dimerization extensively.

1.5. HIV-1 genomic RNA dimerization and the role of NC

The full-length viral RNA can act both as a messenger RNA during translation and genomic RNA which gets packaged into virions during viral assembly (241). The balance between these two functions of genomic RNA (gRNA) is critical for successful viral replication. The expression of viral gRNA is followed by modifications such as 5' capping and polyadenylation which allows for viral gene expression by using the host translational machinery (207). The Gag polyprotein of HIV-1 plays a key role in determining the switch between viral messenger RNA and genomic RNA. Briefly, when viral RNA is first expressed, it is recognized as an mRNA and both 5' cap and IRES-mediated (242) protein translation leads to the expression of Gag and Gag-pol polyproteins. The newly-synthesized Gag molecules can bind the gRNA via the NC region of Gag. NC-gRNA interactions lead to conformational changes in the leader region (121) which simultaneously down-regulates protein translation and increase exposure to structural elements necessary for gRNA dimerization and packaging (243) (Figure 1.12). This process is reviewed in the next sections.

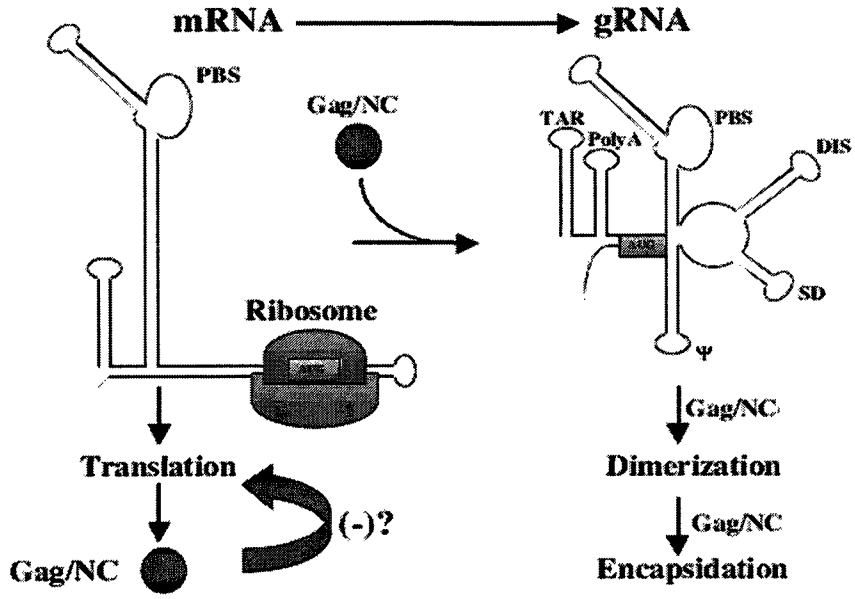


FIGURE 1.13. Role of Gag in determining the faith of viral RNA. The IRES-dependent translation of genomic RNA leads to the expressin of Gag. Gag in turn binds to alternative RNA structures via NC and leads to suppression of viral translation and encapsidation of the dimerized genome (adapted from(121)).

1.5.1. Leader sequence of gRNA

The NC protein specifically recognizes the 5' region of gRNA which consists of a leader region which contains two genetic elements that guide viral translation and packaging. The 5' Internal Ribosome Entry Segment (IRES) can mediate translation of viral proteins by recruiting ribosomal subunits in a cap-independent manner. Also present in the 5' leader region is the packaging signal (Ψ) element responsible for packaging of gRNA into virions (2). Although IRES and the packaging signal share the same sequence, differences in the secondary structures lead to different pathways. The IRES region overlaps with multiple key elements for viral replication. These include the primer binding site (PBS), the dimer initiation site (DIS) and packaging signal (Ψ)(137) (Figure 1.12). It is believed that the 5' leader sequence of HIV-1 can adopt two mutually-exclusive secondary structures. The first is the long distance interactions (LDI) structure and the second is the branched multiple hairpin (BMH). The BMH structure

allows for the recognition of Ψ, DIS and PBS while LDI disrupts DIS secondary structures (185,220). On the other hand, LDI allows for the recognition of the Gag gene start codon while BMH occludes this region (243). The combination of these effects is indicative of distinct roles for these two structures where LDI promotes protein translation from viral mRNA while BMH favors gRNA dimerization and packaging.

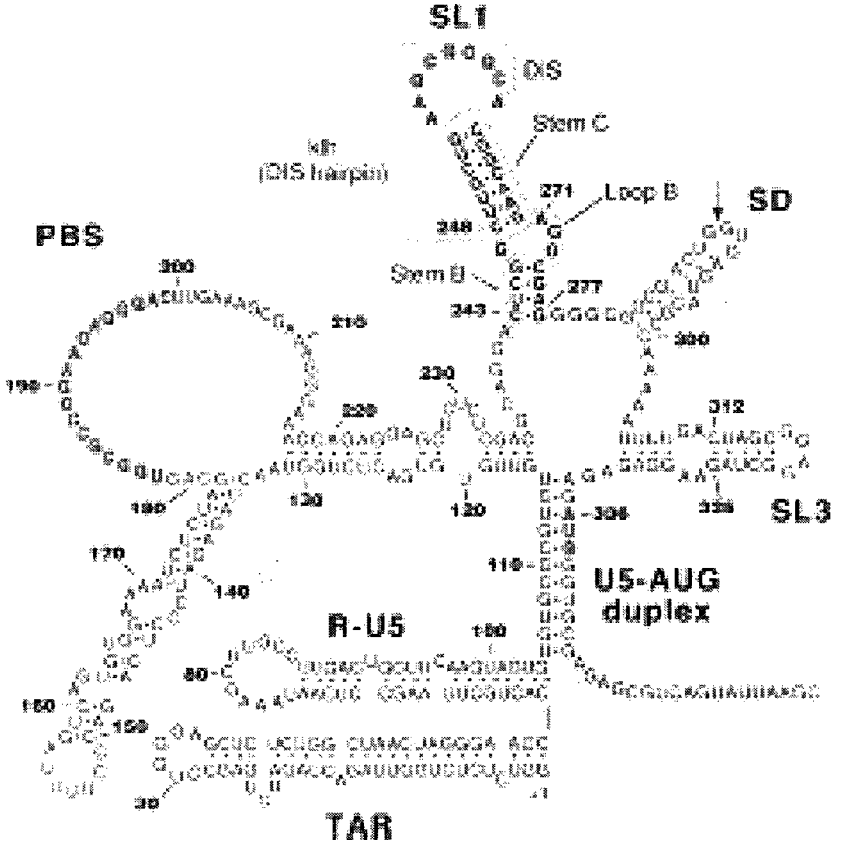


FIGURE 1.14. The secondary structure of the 5' leader sequence of HIV-1 RT (adapted from (153)).

1.5.2. Role of HIV-1 NC in genomic dimerization

Genomic dimerization is an important step in the viral life cycle since impairing this pathway, results in reduced genome packaging as well as defective viral structure (121). DIS sequence dimerization represents one of the major mechanisms of gRNA dimer formation. Intermolecular interactions through the SL1 palindromes (also known as DIS) of two gRNA molecules initially form a "kissing complex". The sequences of the palindromes are GCGCGC or GUGCAC, depending on the HIV subtype (223-225). NC-mediated chaperoning of the gRNA molecules allows for the formation of a stable extended duplex where extensive base-pairing occurs between the two gRNA molecules (137) (Figure 1.15). Introduction of mutations in NC appear to affect dimerization (226). These implicate a role for the zinc fingers as well as the basic linker region of NC (227,228) in genome dimerization. The mechanism through which NC mediates dimerization is still not fully clear since both zinc fingers were mutated at the same time(244) it is not possible to determine the individual role of each finger. Also, the complete deletion of the finger 2 (245) did not result in complete abrogation of dimerization.

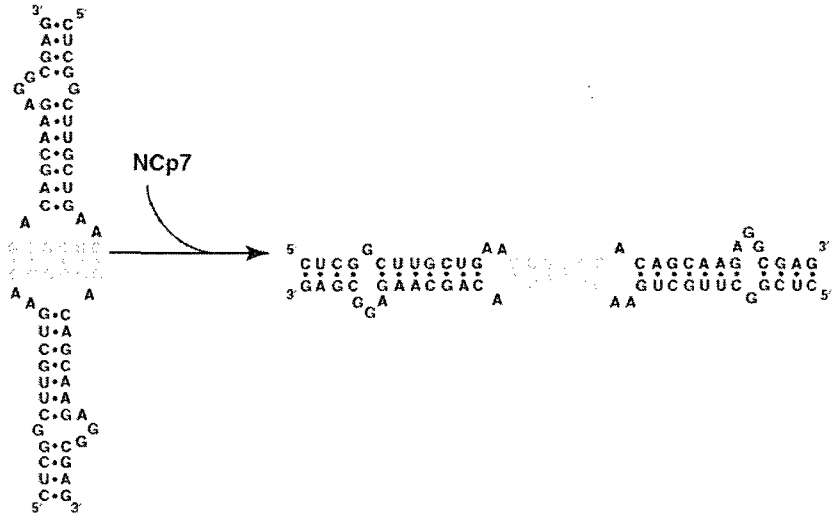


FIGURE 1.15. HIV-1 DIS sequence dimerization. A kissing loop complex can be formed between two SL1 DIS regions. The presence ofHIV-1 NC promotes the formation of the extended duplex leading to gRNA dimerization (121).

Further mutational studies have been performed on the leader region of the genome. Deletion or mutation of SL1 loops, reduces gRNA dimerization (246) supporting the role of DIS and the full SL1 sequence in genome dimerization. However, it has been shown that two SL3 regions can bind to a single NC molecule (222) and that the deletion of this region can also significantly reduce NC-mediated genome dimerization (226,230). This observation implies that gRNA dimerization appears to be achieved through multiple pathways highlighting its important role in viral replication. While many questions remain about the mechanism through which HIV-1 genome dimerization and packaging occurs, the research so far has allowed for a model to be proposed (Figure 1.16). The NC protein (as part of Gag) directly recognizes the SL2 and SL3 regions of Ψ sequence allowing for the selection of the unspliced gRNA for packaging. NC further facilitates the interaction of two gRNA molecules leading to genome dimerization. The multiple NC domains of the Gag polyprotein allow for assembly of multiple Gag molecules. MA is then associated with the plasma membrane of the forming viral particle leading to packaging of Gag along with the dimer genome.

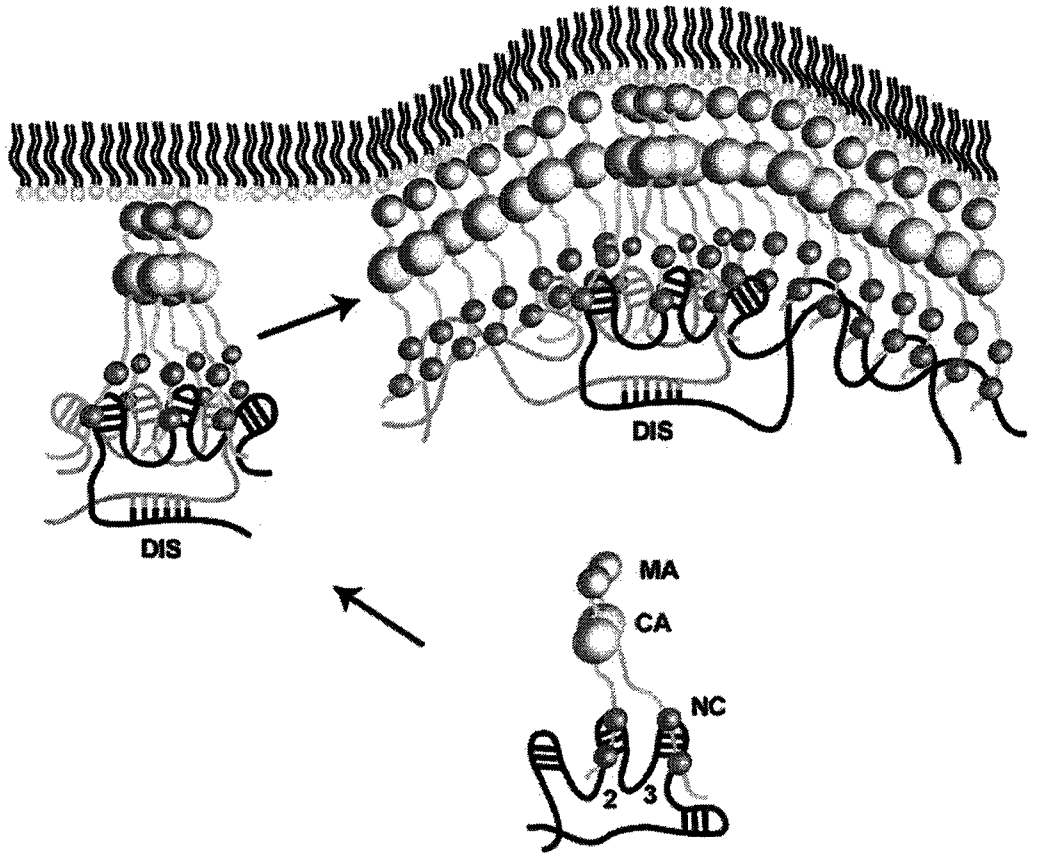


FIGURE 1.16. Model for HIV-1 genome dimerization and packaging (adapted from (145)).

1.6. Conclusion

The current review has highlighted the critical role of NC during the viral life cycle. The importance of this protein for viral survival makes it a good target for antiretroviral therapy. The NC protein of HIV-1 is a dynamic and essential part of the viral life cycle and plays critical roles at various stages of HIV-1 replication such as reverse transcription, integration, genome dimerization and packaging to name a few. However, the exact mechanism through which this protein achieves its many roles still remains to be elucidated. The main focus of this thesis is to investigate the role of HIV-1 NC protein and its amino acid residues on viral genome RNA

dimerization. In the process of our studies we also investigated the effect of our NC-mutated HIV-1 on viral stability, genomic RNA packaging, Pr55gag processing, viral infectivity and RT packaging. We were interested in gRNA dimerization because this absolutely conserved feature in all retroviruses appears essential for viral infectivity via, among others, facilitating viral RNA strand exchange during reverse transcription (247,248). In studying the NC protein, we focused on several factors in designing our experiments and preparing HIV-1 proviral mutants. We wanted to study the electrostatic effect of the basic residues in N-terminus, first and second zinc finger and the linker region in HIV-1 viral cycle with an emphasis on its genomic RNA dimerization (chapter 2). In this section, possible interactions between finger 1 and finger 2 were also studied by several mutants in the flexible linker region. CCHC zinc containing residues and the residues important in NC chaperoning activity shown in *in vitro* experiments were also examined by site directed mutagenesis. In the second step of the study we examined the effect of NC intermediate proteins (NCp15 and NCp9) on viral life cycle again with emphasis on gRNA dimerization. By producing HIV-1 composite viruses and reducing protease activity in these viruses, we studied the effect of Pr55gag protein processing on the gRNA dimerization.

Chapter 2

Mapping of Nucleocapsid Residues Important for HIV-1 Genomic RNA Dimerization and Packaging

Jafar Kafaie, Rujun Song, Levon Abrahamyan, Andew J. Mouland and Michael Laughrea. Mapping of nucleocapsid residues important for HIV-1 genomic RNA dimerization and packaging.Virology. 2008 Jun 5; 375 (2): 592-610. Epub 2008 Mar 17.

Chapter 2 Preface

Similar to other retroviruses, the HIV-1 genome, as in other retroviruses is composed of two non-covalently linked full-length RNA molecules in the form of RNA dimer (249). This dimeric structure appears essential for viral infectivity via, among others, facilitating viral RNA strand transfer during reverse transcription (30,248). Newly released protease inactive HIV-1 viruses lack any dimeric RNA which then accumulates with time (153). One or more maturation products of the viral polyproteins are required for fast or complete formation of thermolabile viral RNA dimers, and for refolding of thermolabile dimers to mature and thermostable dimmers (153). Before starting my project, there were some papers published about the role of NC protein in retroviral genomic RNA dimerization. The earliest work done on retroviruses other than HIV-1 (250) showed that an mutation in the first zinc finger of Rous Sarcoma Virus diminished its genomic RNA dimerization by 50 %. Regarding HIV-1, *in vitro* work implying the NC protein in dimerization of partial RNA transcripts started with Muriaux et al. in 1996 (251). The first and only *in vivo* evidence for a role of NC in genomic RNA dimerization was published by Laughrea et al. in 2001 (252), who showed that a mutation in the second zinc finger of the NL4-3 strain of HIV-1 can decrease genomic RNA dimerization as much as deleting its DIS. Many unanswered questions still remain to be addressed, like the role of finger domains of NC, the hydrophobic plateau, the conserved Glycines, or zinc containing residues, in HIV-1 genomic RNA dimerization. Moreover, there were no *in vivo* data regarding any possible interaction between two fingers of NC protein in chaperoning its genomic RNA dimerization. Though the role of the linker and N-terminus region in gRNA dimerization was studied to some extent in 2001 (252), there was a need for a deeper investigation of the role of these segments, especially their electrical charge and their steric effect in gRNA dimerization.

In an effort to answer some of these questions, we started a systematic study of the effect of HIV-1 NC mutations on gRNA dimerization by site directed mutagenesis. We were careful to choose and study those residues that are conserved, have a positive electric charge or are believed to be important in chaperoning property of NC. We were also interested in zinc containing residues and mutations that we expected to distort the shape of the hydrophobic plateau derived from both zinc fingers.

In chapter 2 of my thesis I will show the results of a systematic study of the role of NC components in gRNA dimerization, viral infectivity, virus stability, RT packaging, Pr55gag processing and finally, viral gRNA packaging. This will involve the use of 40 mutants. 38 of these mutants were studied for the first time, and substantially increase our knowledge on the role of NC in viral gRNA dimerization.

2.1. Abstract

Retroviral genomic RNA (gRNA) dimerization appears essential for viral infectivity, and the nucleocapsid protein (NC) of human immunodeficiency virus type 1 (HIV-1) facilitates HIV-1 gRNA dimerization. To identify the relevant and dispensable positions of NC, 34 of its 55 residues were mutated, individually or in small groups, in a panel of 40 HIV-1 mutants prepared by site-directed mutagenesis. It was found that the amino-terminus, the proximal zinc finger, the linker, and the distal zinc finger of NC each contributed roughly equally to efficient HIV-1 gRNA dimerization. The N-terminal and linker segments appeared to play predominantly electrostatic and steric roles, respectively. Mutating the hydrophobic patch of either zinc finger, or substituting alanines for their glycine doublet, was as disabling as deleting the corresponding finger. Replacing the **Cys**X_2**Cys**X_4**His**X_4**Cys** motif of either finger by **Cys**X_2**Cys**X_4**Cys**X_4**Cys** or **Cys**X_2**Cys**X_4**His**X_4**His**, interchanging the zinc fingers or, replacing one zinc finger by a copy of the other one, had generally intermediate effects; among these mutations, the His23->Cys substitution in the N-terminal zinc finger had the mildest effect. The charge of NC could be increased or decreased by up to 18 %, that of the linker could be reduced by 75 % or increased by 50 %, and one or two electric charges could be added or subtracted from either zinc finger, without affecting gRNA dimerization. Shortening, lengthening, or making hydrophobic the linker was as disabling as deleting the N-terminal or the C-terminal zinc finger, but a neutral and polar linker was innocuous. The present work multiplies by 4 and by 33 the number of retroviral and lentiviral NC mutations known to inhibit gRNA dimerization, respectively. It shows the first evidence that gRNA dimerization can be inhibited by: 1) mutations in the N-terminus or the linker of retroviral NC; 2) mutations in the proximal zinc finger of lentiviral NC; 3) mutations in the hydrophobic patch or the conserved glycines of the proximal or the distal retroviral zinc

finger. Some NC mutations impaired gRNA dimerization more than mutations inactivating the viral protease, indicating that gRNA dimerization may be stimulated by the NC component of the Gag polyprotein. Most, but not all, mutations inhibited gRNA packaging; some had a strong effect on virus assembly or stability.

2.2. Introduction

The genome of HIV-1, a member of the *Retroviridae* family, is a single-stranded RNA. Two copies are packaged and they can dimerize. This dimeric structure appears essential for viral infectivity via, among others, facilitating viral RNA strand exchange during reverse transcription (247,248). It is not clear that protein-free formation of viral genomic RNA (gRNA) dimers can occur in the infected cell, the test tube, or even the virus. No gRNA dimers are detectable in newly released protease inactive (PR^-) HIV-1, but thermolabile and partially dimeric viral RNA dimers slowly accumulate with time (253). One or more maturation products of the viral polyproteins is required for fast or complete formation of thermolabile viral RNA dimers, and for refolding of thermolabile dimers into mature, thermostable dimers (253). Among these maturation products, a central role is attributed to the HIV-1 nucleocapsid protein (NC) for at least four reasons: 1) avian and murine NC are implicated in the dimerization of avian and murine gRNA (118,250,254-256). 2) NC stimulates the *in vitro* dimerization of partial HIV-1 RNA transcripts containing the gRNA dimerization initiation site (131,146,251). 3) the HIV-1 NC stabilizes HIV-1 RNA dimers (251) and murine retroviral RNA dimers (146); 4) a mutation in NC can decrease HIV-1 gRNA dimerization as much as disruption of the dimerization initiation site (252). Disruption of the dimerization initiation site (DIS) reduces by $\geq$ 50 % the proportion of viral RNA dimers in isolated viruses (246,252,253,257-260). Unless otherwise noted, NC designates the HIV-1 NC, and RNA designates HIV-1 RNA.

The DIS is located in the apical loop of SL1, a 35 nucleotide-long stem-loop belonging to the 5' untranslated region of gRNA. This has stimulated studies on the role of NC in the dimerization of SL1 or molecules similar to SL1, with emphasis on: 1) the rearrangement of these molecules from a metastable kissing dimer to a stable extended duplex (261-266); and 2) the effect of NC mutations on the dimerization of a 39mer RNA similar to wild type (WT) SL1 but lengthened by two additional base-pairs for technical reasons (244,245). A role for the zinc fingers and the basic regions of NC was identified, but there were two major ambiguities: 1) substituting completely scrambled N-terminal or linker sequences for the WT sequence did not inhibit dimerization, as if nonspecific effects were studied under the in vitro conditions (245); 2) the two zinc fingers were jointly mutated, preventing an identification of the implicated finger(s) (244).

The NC protein is a sequence of 55 amino acid residues, 15 of which are highly basic and only four highly acidic. Like all lentiviral NC, it possesses two zinc containing motifs (often called zinc fingers) of the form **Cys**X_2**Cys**X_4**His**X_4**Cys**, where X = variable or conservatively substituted amino acid residue (267-269) Zinc finger 1 (residues 14 to 28) and zinc finger 2 (residues 36 to 49) are preceded by a N-terminal segment (residues 1 to 13, containing four arginine and lysine residues and no acidic residues), separated by a linker peptide (residues 29 to 35, containing four arginine and lysine residues and no acidic residues), and followed by a short C-terminal portion (Table 1). NC is initially part of the precursor polyproteins Pr55gag and Pr160gag-pol. As proteolytic maturation of Pr55gag progresses under the direction of the viral protease, NC becomes part of progressively smaller proteins called NCp15 (NCp7-p1-p6), NCp9 (NCp7-p1) and NCp7, the numbers indicating the approximate molecular weights in thousands of daltons. These four proteins account for 95% of NCs. The other NCs are part of Pr160gag-pol and maturation products thereof, which are NCp120

(NCp7 linked to the Pol polyprotein), NCp8 (NCp7 + an 8 residue Pol coded peptide), and possibly some NCp7 (143,144). In mature virions, almost all NC is in the form of NCp7.

In vitro studies of the dimerization of partial HIV-1 RNA transcripts can be prone to artefacts (270). It is thus crucial to also investigate the effect of mutations on the properties of WT gRNA produced by viruses. Regarding the effect of NC mutations, gRNA dimerization was unaffected by a 75 % reduction of the number of basic residues in the N-terminus or the linker; and inactivating the C-terminal zinc finger inhibited gRNA dimerization as much as inactivating the DIS, but clearly less than inactivating the viral protease (252). Since mature NC is believed to be solely responsible for the increased dimerization seen in WT virions relative to PR^- virions, this differential effect suggests that NC segments other than the C-terminal zinc finger can modulate gRNA dimerization. Thus many issues were left unresolved, such as the role of the N-terminal zinc finger and its components (e.g. its hydrophobic patch (or cleft), its highly basic residues, its highly conserved histidine and glycines), the role of the corresponding components in the distal zinc finger, a deeper inquiry into whether the N-terminal and linker segments can modulate gRNA dimerization. Moreover, the NC mutations were studied in the NL4-3 isolate (252). HIV-1_{NL4-3} is less impaired by the neutralization of basic NC residues than the HXB2-BH10 isolate (271,272), even though it is also a subtype B strain (HIV-1_{NL4-3} has a NY5 5' half and a HXB2-Lai 3' half)(273).

Here, the impact of 40 NC mutations on gRNA dimerization was investigated. Thirty-eight were investigated for the first time; the remaining two were previously studied in HIV-1_{NL4-3} (252). The 40 mutations were engineered in the first four segments of NC from HIV-1_{HXB2}. They collectively involve 34 of their 49 residues (the six-residue C-terminal segment was not mutated). In each investigated segment, relevant and dispensable residues were identified. The results reveal that the amino-terminus, the proximal zinc finger, the linker, and the distal zinc finger of NC each

contributed roughly equally to efficient gRNA dimerization, the N-terminal and linker segments playing predominantly electrostatic and steric roles, respectively. Mutating the hydrophobic patch or the conserved glycines of either zinc finger was as disabling as deleting the corresponding finger; some mutations involving as little as three NC residues impaired gRNA dimerization more than inactivating the viral protease; the charge of NC could be increased or decreased by up to 18% without affecting gRNA dimerization.

NC mutations can also impair other late infection events such as gRNA packaging (51,124,130,239,274-277), virus assembly or stability (109,192,201,222,276,278-280), as well as packaging of Pr160gag-pol or reverse transcriptase activity (124,221). We investigated the effect of our NC mutations on these functions, as well as on viral replication.

2.3. Results

HeLa cells were transfected in parallel with equal amounts of pSVC21.BH10 or mutant proviral vectors. Proviral vector pSVC21.BH10 encodes an infectious HIV-1_{HXB2} molecular clone derived from the IIIB strain of HIV-1 (246). After 48 h, viruses were isolated from the culture supernatant, their capsid protein (CA) and reverse transcriptase (RT) content was measured, and their gRNA was extracted, electrophoresed on a non-denaturing agarose gel and visualized by Northern blotting with a ^{35}S-labeled HIV-1 riboprobe, followed by autoradiography. Prior to virus purification, a small volume of culture supernatant was kept to measure its CA content and determine viral replication per unit of supernatant CA (Materials & Methods). The identity of the various mutants is described in Table 1. For comparative purposes, the gRNA dimerization and Pr55gag processing seen in protease-inactive (PR^-) virions was also included. The aspartic acid at position 25 of the viral protease active site was replaced by arginine in PR^- virions; gRNA dimerization had previously been investigated in this PR^- context (253).

Mutations in one NC segment only are described first. Mutations in two NC segments are presented in a separate section (called "joint mutations...") at the end. For each NC mutant, viral replication, gRNA dimerization, gRNA packaging, RT packaging, virus stability, and Pr55gag polyprotein processing was typically measured. Genomic RNA packaging was measured by dot blot hybridization. RT packaging was defined as CA-normalized RT activity: the RT activity of isolated virions divided by their CA content, relative to the ratio found in WT samples. Unprocessed Gag-Pol, though it contains RT, is not recorded by the RT packaging assay. Virus stability was defined as CA content of purified viruses divided by CA content of the culture supernatant, relative to the ratio found in WT samples (Materials & Methods).

	Construct name	N-terminus 1	Zinc finger 1 14	Linker 29	Zinc finger 2 36	C-terminus 50
1	HXB2	MQRGNFRNQRKIV	KCFNCGKEGHTARNC	RAPRKKG	CWKCGKEGHQMKDC	TERQAN
2	R7	——S—SS—				
3	R7E	——E—EE—				
4	N+	—K—K—				
5	3AF1		A——A——A—			
6	3EF1		E——E——E—			
7	N17K		—K—			
8	F16A		—A—			
9	2GAF1		—A—A—			
10	H23A		——A—			
11	H23C		——C—			
12	C28H		——H			
13	C28S		——S			
14	NC2-2		—WK——QMKD—			
15	ΔF1		G G			
16	S3			—SSS—		
17	S3E			—EEE—		
18	FVI			—FV I—		
19	P31A			—A—		
20	ΔAP			--		
21	LL			——GGGGG		
22	ΔLinker			G G		
23	L+			—KK—		
24	2KAF2				——A——A—	
25	2KEF2				——E——E—	
26	2GAF2				—A—A—	
27	W37A				—A—	
28	W37F				—F—	
29	K38N				—N—	
30	H44A				——A—	
31	H44C				——C—	
32	C36S				S—	
33	C49H				——H	
34	NC1-1				—FN——TARN—	
35	ΔF2				G G	
36	NL	——S—SS—		—FV I—		
37	FL		——A—	—FV I—		
38	FF		——A—		—A—A—	
39	NC2-1		—WK——QMKD—		—FN——TARN—	
40	CH		——C—		——H	
41	HC		——H		——C—	
42	HXB2	MQRGNFRNQRKIV	KCFNCGKEGHTARNC	RAPRKKG	CWKCGKEGHQMKDC	TERQAN
		1	14	29	36	50

Table 2.1. The wild-type sequence is indicated at the top and bottom lines; using a ruler joining the two HXB2s, mutated residues can be easily identified. The sequence is dividedinto segments and the residue number corresponding to the first residue of each segment is given. One letter abbreviations are used for each amino acid residue. Lines indicate amino acid residues that are the same as wild type. Deletions are shown by blank spaces in the mutant sequences; blank spaces within the wild-type sequence are present only to simplify presentation of the mutants. Insertions are indicated by letters where there is a blank in the wild-type sequence.

Effect of mutations in the N-terminus

Three mutations were studied. Mutation R7 (Arg7 and ArgLys11 replaced by serines) blocked viral replication and inhibited each specific function studied, while R7E (Arg7 and ArgLys11 replaced by glutamic acids) further reduced gRNA and RT packaging per unit of capsid protein.

Mutations R7 and R7E each reduced gRNA dimerization to 77-79% of the WT level (Fig. 1 and Table 2; $P < 0.01$). This represents the first evidence that mutations in the N-terminus of retroviral NC can inhibit retroviral gRNA dimerization. Mutation N+ (Asn5 and Gln9 replaced by lysines) had no effect on gRNA dimerization and RT packaging (Fig. 1, Table 2); but it reduced gRNA packaging and enhanced virus stability (Table 2). The N+ supernatant appeared almost as infectious as WT (Fig. 2A), as if the enhanced virus stability had partially compensated for the packaging defect. While R7E strongly reduced gRNA packaging, comparison of N+ with R7 suggests that the gain of two positive charges in the N-terminus is more detrimental to gRNA packaging than the loss of three positive charges (Table 2).

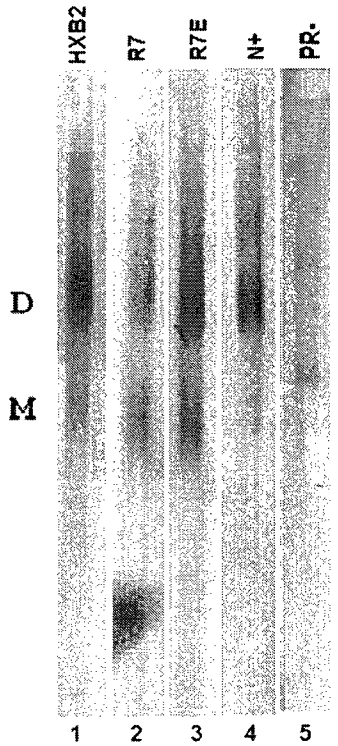


Fig. 2.1. Dimerization level of viral RNA isolated from HIV-1_{HXB2} and virions mutated in the N-terminus segment of the nucleocapsid protein. Genomic RNAs extracted from the respective virions were electrophoresed on a 1% non-denaturing agarose gel and analyzed by Northern blotting. The representative lanes contain viral gRNA isolated from one 35 mm tissue culture dish, and the autoradiographic exposure times varied from 15 min to 9 h. Mutants are defined in Table 1. D: dimer. M: monomer. BH10 (HXB2) gRNA samples were 77 % ± 0.5 % dimeric (n = 29) and PR^- gRNA samples were 46 % ± 2.5 % dimeric (n = 3). The gRNA dimerization level is independent of the amount of gRNA electrophoresed or of the concentration of DNA used in transfections (25-fold range of gRNA/proviral DNA concentrations tested with BH10 and an HIV-1_{HXB2} mutant bearing an inactivated dimerization initiations site (Song *et al.*, 2007)) (not shown). It is also known from a previous study that highly reduced HIV-1 gRNA packaging caused by NC mutations need not impair gRNA dimerization (Laughrea *et al.*, 2001).

Confirming previous data also obtained with HIV-1_{HXB2} produced by HeLa cells (253,256), the gRNA isolated from PR^- virions was 46% dimeric, i.e. 60% of WT (Fig. 1 and Table 2). Therefore a mutation in NC could a priori inhibit gRNA dimerization via impairing proteolytic maturation of the HIV-1 Gag precursor polyproteins. The proteolytic processing of Pr55gag occurs in several steps denoted primary, secondary, and tertiary cleavage events. The primary cleavage site is located between spacer peptide p2 and NC, and yields a NCp15 component and a MA-CA-p2

component of approximately 41 kilodaltons in molecular weight. The secondary cleavage sites are located between the matrix protein (MA) and CA, to yield MA and CA-p2, and between spacer peptide p1 and protein p6, to yield NCp7-p1 and p6 (50). Pr55gag maturation was estimated by extracting proteins from purified viruses, separating them by SDS gel electrophoresis and analysing them by Western blotting using an anti-capsid antibody (Materials & Methods). This gives access to the yield of the cleavage between p2 and NC, MA and CA, and CA and p2. The absence of cleavage between NCp7 and p1 or between p1 and p6 had respectively no and little effect on gRNA dimerization (Kafaie et al., in preparation). Therefore the key proteolytic blockages susceptible to inhibit gRNA dimerization were accessible in our Western blots.

Fig. 2

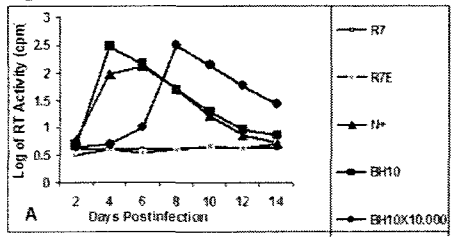


Fig. 2 A

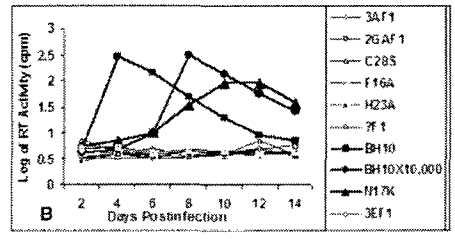


Fig. 2 B

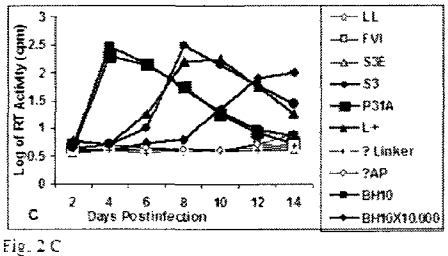


Fig. 2 C

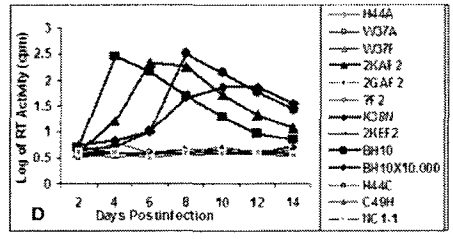


Fig. 2 D

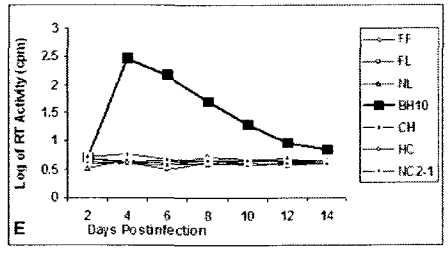


Fig. 2 E

Fig. 2.2. Viral replication of HIV-1_{HXB2} mutated in the N-terminus (A), the N-terminal zinc finger (B), the linker (C), the C-terminal zinc finger (D) of the nucleocapsid protein, or mutated in two separate segments of the protein (E). MT2 cells were infected with an amount of undiluted progeny virus equal to 10 ng of CAp24 antigen. Virus growth was monitored by measuring reverse transcriptase activity (cpm/μl) in culture fluids at various times. The replication of 10,000X diluted wild-type HXB2 was also studied for comparative purposes. Mutants are defined in Table 1.

The reduced gRNA dimerization in mutant R7 is unlikely to be due to poor Pr55gag processing because Pr55gag was 73 % processed (Fig. 3A). Interestingly, Pr55gag was only 60 % processed in R7E without any further reduction in gRNA dimerization relative to R7 (Fig. 3A; Table 2). This suggests that a limited degree of incomplete Pr55gag processing has no impact on gRNA dimerization, or that replacing basic residues by acidic residues has less direct influence on dimerization than replacing them by neutral ones. The first possibility seems more likely: presumably, full gRNA dimerization requires no more, and possibly much less, than one or two thousand processed Pr55gag per virus, out of the ~ 4 000 incorporated in each particle (184).

Fig. 3

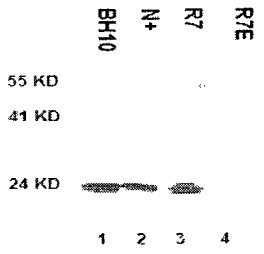


Fig. 3 A

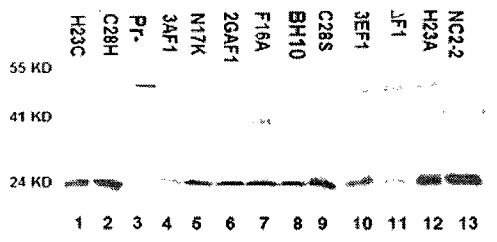


Fig. 3 B

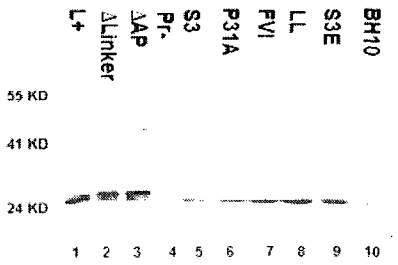


Fig. 3 C

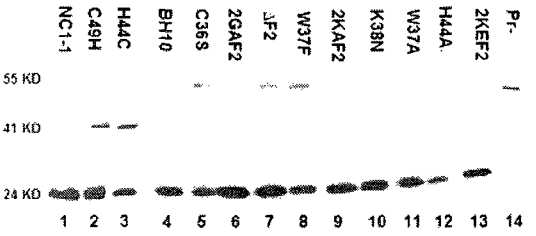


Fig. 3 D

Fig. 3 E

Fig. 2.3. Pr55gag maturation in HIV-1_{HXB2} mutated in the N-terminus (A), the N-terminal zinc finger (B), the linker (C), the C-terminal zinc finger (D) of the nucleocapsid protein, or mutated in two separate segments of the protein (E). Proteins extracted from purified viruses were assessed by SDS-polyacrylamide gel electrophoresis, followed by visualization by Western blotting using capsid-reactive antibodies (Materials and Methods). Pr55gag runs at the 55 KD position. After cleavage of Pr55gag at the primary cleavage site, and before cleavage at the secondary and tertiary sites, the capsid is part of a matrix-capsid oligoprotein running at the 41 KD position. Mutants are defined in Table 1.

Effect of mutations in the zinc fingers

Depending on authors, zinc finger 1 extends from Val13, Arg14 or Cys15 to Cys28, Arg29 or Ala30, while zinc finger 2 extends from Gly35 or Cys36 to Cys49, Thr50 or Glu51; finger 1 is usually described as one residue longer than finger 2 or identical in length (106,145). To simplify presentation, mutations 3AF1 and 3EF1 were classified as within zinc finger 1 (Table 1). This acknowledges the more basic environment of the first CCHC motif (Cys15 to Cys28): it is sandwiched between two highly basic sequences, whereas the second CCHC motif (Cys36 to Cys49) is followed by six to nine residues (nine in NCp9) of net charge 0.

Zinc finger 1

Eleven mutations were studied. All mutations blocked viral replication, except N17K (Asn17 -> Lys), which was about 10 000 times less infectious than WT (Fig. 2 B). Thus single, double and triple point mutations in zinc finger 1, no matter their position (residues 14, 16, 17, 19, 20, 22, 23 or 28) can have devastating effects on viral replication.

Dimerization

Nine mutations substantially hindered gRNA dimerization; among the remaining two mutations, N17K had no effect, and H23C had little effect (Fig. 4, lanes 6 and 12; and Table 2). Deletion of zinc finger 1 (ΔF1) inhibited gRNA dimerization to approximately 72% of WT, i.e. somewhat less than protease inactivation (Fig. 4, lane 3, and Table 2). At least four mutations inhibited gRNA dimerization as much as deleting the whole finger. They include two of the six point mutations studied (F16A and H23A), the substitution of alanine for the two glycines (2GAF1), and of glutamic acid for the two lysines and the arginine (3EF1) (Fig. 4, lanes 2, 3, 5, 7 and 9; Table 2). F stands for phenylalanine, except in F1 and F2, where it stands for zinc finger. The gRNA from

mutants 3AF1, C28H and C28S was somewhat more dimeric than H23A and 3EF1 (Fig. 4; Table 2). Thus mutations C28H and C28S inhibited gRNA dimerization to a level intermediate between the higher level of H23A and the lower level of H23C (Fig. 4, lanes 7-8, 11-12; Table 2; P = 0.05) (full appreciation of these intermediate differences requires densitometric analysis of several gels from independent experiments). The effects of C28H and C28S on the other measured parameters were also quite similar (Table 2). The C28H and C28S results suggest that tetrahedral zinc coordination does not overwhelmingly matter. For example, other steric consequences of H23A, C28H or C28S might have more influence. Interestingly, the proximal zinc finger of C28H NC coordinates cobalt tetrahedrally, and with an affinity comparable to WT, but adopts two interverting folded conformations, each differing from the native conformation and modifying the orientation of Phe16 (281). C28S NCp7 binds zinc with an affinity > 40-fold lower than WT (282). This affinity could therefore be higher than or similar to that of various host zinc binding proteins (134). Natural CCHH zinc fingers belong to double-stranded DNA binding proteins and are at least seven amino acid residues longer than retroviral zinc fingers (283-285).

Mutation NC2-2 replaces zinc finger 1 by a second copy of finger 2. It was less disabling than small mutations such as H23A, 2GAF1 or F16A (Fig. 4, lanes 5, 7, 9, 10 Table 2). Thus jointly mutating the six distinctive residues of zinc finger 1 into related residues impaired gRNA dimerization less than mutating one or two of its conserved residues into unrelated residues; NC2-2 NC might be more WT-like in shape and/or chemical properties than H23A, 2GAF1 or F16A NC. On the other hand, the RNA dimer from NC2-2 virions exhibited a heterogeneous migration profile, as if the NC2-2 gRNAs were more conformationally diverse than the H23A, 2GAF1 or F16A gRNAs. Heterogeneous dimer migration was seen previously as a result of deleting the DIS (253), most of SL1 (Hill *et al.*, 2003) or blocking synthesis of Pr160gag-pol (286). In Rous sarcoma virus,

the gRNA dimerization level was reduced by mutations in zinc finger 1 (118,250,254,255). Our results extend the phenomenon to lentiviruses.

Other than 3EF1, F16A and ΔF1, the mutations had no or few effects on Pr55gag proteolytic processing (Fig. 3 B). F16A did not impair the primary p2-NC cleavage, but moderately inhibited the secondary MA-CA cleavage site (Fig. 3 B, lane 7). Note that ΔF1 inhibited Pr55gag processing much more than 2GAF1 and H23A (Fig. 3 B, lanes 6, 11 and 12; Table 2), without further impairing gRNA dimerization. This suggests that a 50% reduction in Pr55gag processing has no impact on gRNA dimerization. Analogously, no gRNA packaging defect was seen virions containing 80% inactive NC (80% of their NCs had the double mutation Cys15Tyr + Cys18Tyr) (75).

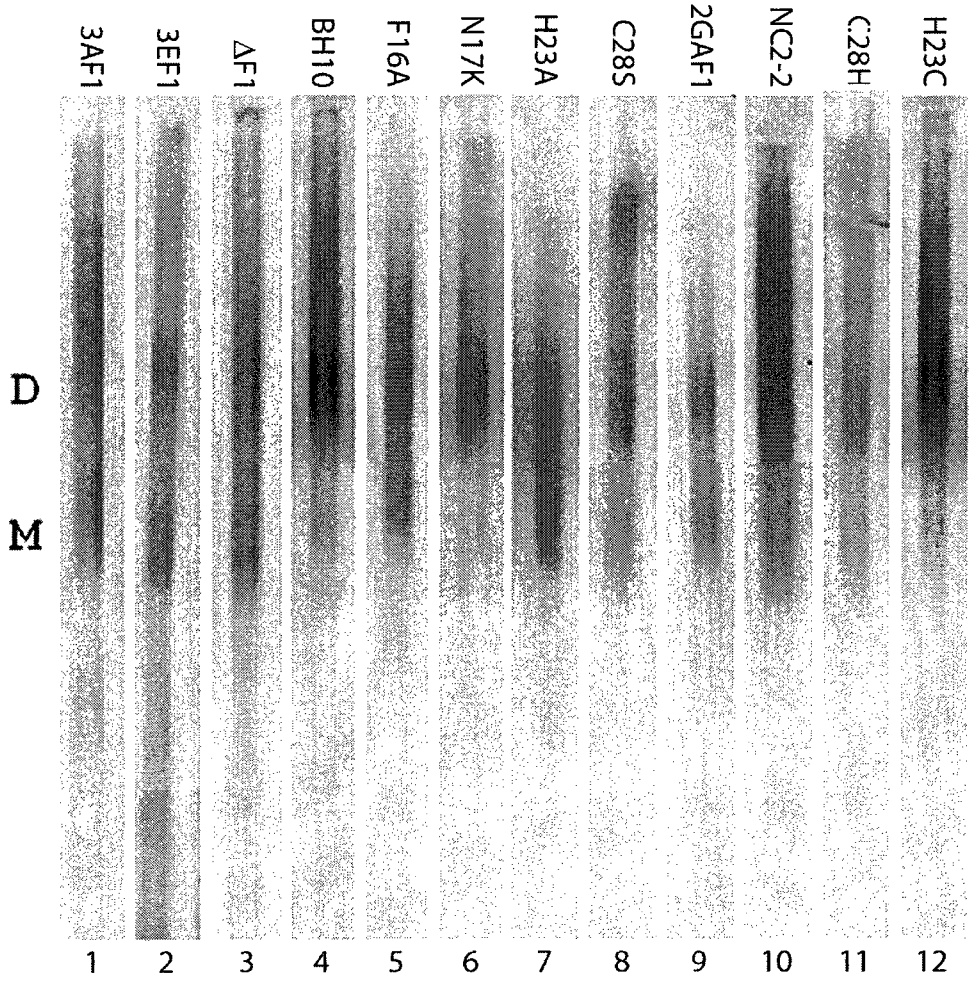


Fig. 2.4. Dimerization level of viral RNA isolated from HIV-1_{HXB2} mutated in the N-terminal zinc finger of the nucleocapsid protein. Experimental conditions as in Fig. 1. Exposure times varied from 15 min to 14 h.

Packaging

Most mutations inhibited gRNA packaging as much, or nearly as much, as deleting the whole finger. F16A and 2GAF1 appeared particularly inhibitory. One striking exception was N17K: it stimulated gRNA packaging 2 to 3-fold relative to WT (Table 2).

Virus stability was reduced ≥ 5-fold by mutation 3EF1. Most of the other mutations, including ΔF1, had approximately a 2-fold inhibitory effect. RT packaging was little affected by

mutations H23C and C28H, and reduced 1.5 to 3-fold, depending on the mutation, by the other mutations (Table 2).

Region		Construct name[a]	Viral replication[b]	gRNA dimerization[c]	gRNA packaging[c]	RT packaging[c]	Virus stability[c]	Pr55gag processing[d]
	1	HXB2	+	100	100	100	100	97
	2	PR−	−	60±3	nd	nd	nd	0
N-Terminus	3	R7	−	77±2	48±2	nd	46±5	73
	4	R7E	−	79±5	8±2	19±2	32±5	60
	5	N+	+	102±3	23±4	91±5	239±33	100
Zinc finger 1	6	3AF1	−	79±1	42±2	28±4	nd	98
	7	3EF1	−	68±4	22±3	24±4	15±2	80
	8	N17K	+/−	97.5±2	250±73	59±4	69±7	97
	9	F16A	−	76±2	15±2	nd	nd	77
	10	2GAF1	−	73±3	18±2	nd	56±6	91
	11	H23A	−	69±4	41±3	39±6	36±4	89
	12	H23C	−	91±1.5	30±2	71±2	25±4	98
	13	C28H	−	81±1.5	36±11	82±3	nd	99
	14	C28S	−	81±3.5	50±5	60±5	39±12	96
	15	NC1-2	−	86±2	34±3	44±3.5	nd	93
	16	ΔF1	−	72.5±3	25±10	nd	43±7	54
Linker	17	S3	−/+	94±3.5	39±13	nd	97±14	98
	18	S3R	−	45±2	28±5	10±3	24±4	64
	19	PV1	−	74±2	40±6	nd	nd	97
	20	P11A	+	96.5±2.5	31±4	nd	84±10	98
	21	ΔAP	−	75±2	34±4	95±11	12±3	90
	22	LL	−	72±3	65±4	45±5	42±14	96
	23	ΔLinker	−	75±2	16±6	nd	21±4	89
	24	L+	+/−	96±3	20±7	nd	100±15	98
Zinc finger 2	25	2KAF2	+/−	94±2	50±2	nd	68±8	95
	26	2KEF2	−	75±2	14±4	nd	14±2	96
	27	2GAF2	−	65±2	17±6	88±8.5	69±9	98
	28	W37A	−	71±2	37±5	nd	96±8	100
	29	W37F	−	76±4	30±7	71±5	44±5	71
	30	K38N	+/−	101±4	109±17	nd	18±5	95
	31	H44A	−	70±2	32±3	88±8.5	33±11	97
	32	H44C	−	77±4.5	40±10	85±4.5	30±4	74
	33	C46S	−	92±3	17±6	nd	nd	81
	34	C49H	−	83±2	21±4	90±9.5	27±4	78
	35	NC1-2	−	84±1	83±5	57±7	40±9	96
	36	ΔF2	−	76±1.5	27±2	89±8	26±4	84
Joint mutations	37	NL	−	79.5±3	21±1	22±3	48±4	70
	38	FL	−	46±3.5	13±4	nd	48±4	65
	39	FF	−	50±5	13±6	nd	nd	69
	40	NC2-1	−	82±1.5	98±14	60±5	42.5±7	82
	41	CH	−	78±10	15±10	23±4	19±5	76
	42	HC	−	61±1.5	20±12	28±6	22±4	82

Table 2.2. Except for the analysis of Pr55gag processing, the values for HXB2 are arbitrarily set at 100, and the values for the mutants are expressed as % of wild-type level.Genomic RNA dimerization and Pr55gag processing numbers were obtained by densitometric analysis. nd: not done, or less than 2 independent experiments.

[a] Mutants are defined in Table 2.1, except for PR−, which stands for protease-inactive virions: the viral protease active site aspartic acid residue at position 25 was replaced by arginine (253).

[b] +: identical or close to wild type; +/−: equivalent to wild type diluted 100 to 10 000 times; −/+: equivalent to wild type diluted more than 10,000 times; −: no viral replicaton detected.

[c] Margins of error designate the standard error of 3 to 10 independent experiments for gRNA dimerization (5 on average), 2 to 5 independent experiments, for gRNA packaging (3.5 on average), and 2 independent expriments for RT packaging and virus stability. Genomic RNA packaging was measured by dot–blot hybridization (Materials and methods).

[d] Numbers describe the percentage of the Pr55gag that was processed to the CAp24 level, as seen in Fig. 2.3. This percentage is, in several mutants, lower than the percentage of Pr55gag that was proteolytically cleaved. For example, ten mutants (e.g. F16A, 2GAF1, NC2-2, etc...) contained as little Pr55gag as HIV-1HXB2, but less CAp24 than HIV-1HXB2, usually because of an accumulation of MA-CA processing intermediates.

Zinc finger 2

Twelve mutations were studied (Table 1). Each mutation blocked viral replication, except 2KAF2 (substitution of alanine for Lys41 and Lys47) and K38N, which were about 100 times and 10 000 times less infectious than WT, respectively (Fig. 2 D). Thus point mutations W37A, H44A, H44C, C49H or even W37F (W standing for tryptophan) had devastating effects on viral replication, even though W37F simply replaced an aromatic residue by another that was wild-type in zinc finger 1.

Dimerization

Nine mutations considerably inhibited gRNA dimerization, while C36S, K38N and 2KAF2 had no or little effect (Fig. 5, lanes 2, 3, and 10). K38N and 2KAF2 indicate that neutralising the positively charged residues of zinc finger 2 had no or little impact on gRNA dimerization. Deleting zinc finger 2 (ΔF2) had approximately the same impact as deleting zinc finger 1 (compare lane 8 of Fig. 5 with lane 3 of Fig. 4). Mutations 2KEF2 (replacing Lys41 and Lys47 by glutamic acids), 2GAF2 (replacing the two glycines by alanines), W37A, W37F, H44A, and H44C, which include four of the seven point mutations studied, reduced gRNA dimerization as strongly as deleting zinc finger 1 or zinc finger 2 (Fig. 5; Table 2). Mutations 2GAF2 and W37F are striking because both were disabling and highly conservative. Mutations C49H and NC1-1 (replacing zinc finger 2 by a second copy of the N-terminal finger) had intermediate effects (larger than C36S and 2KAF2, but smaller than 2KEF2, 2GAF2, W37A, H44A, and ΔF2 [Fig. 5, lanes 11 and 13; Table 2; $P < 0.05$; see

below]). Mutation NC1-1 was, on average, less impairing than W37F (Table 2; P = 0.07), even though NC1-1 includes W37F plus five other mutations (Table 1). Thus phenylalanine at position 37 may be more acceptable when the surrounding residues are adapted. Recalling NC2-2, the RNA dimer band from NC1-1 had a more heterogeneous migration profile than in WT or W37F. Most mutations in zinc finger 2 had effects similar to the corresponding mutations in zinc finger 1: compare 2KEF2 with 3EF1, 2GAF2 with 2GAF1, W37A with F16A, K38N with N17K, H44A with H23A, C49H with C28H, and ΔF2 with ΔF1. But H44C was much more inhibitory than H23C, as if a distal CCCC motif had more difficulty coordinating zinc than a proximal CCCC motif. While the affinities for zinc of the H23C, H44A, and C36S zinc fingers appear reduced by 4 to 12-fold (282), 100-fold, and 10 000-fold (134), respectively, that of the H44C zinc finger is unknown.

The reduced gRNA dimerization is unlikely to result from insufficient Pr55gag maturation because the processing yield was ≥ 75% of WT for all mutants (Fig. 3D). The W37A, W37F and F16A results represent the first evidence that impairing the hydrophobic patch of a retroviral zinc finger can inhibit gRNA dimerization. W37A is particularly unambiguous as it had no impact on Pr55gag processing. In vitro experiments had previously shown that F16A and W37A impaired the stabilization of a short dimeric transcript from a murine retroviral RNA (91).

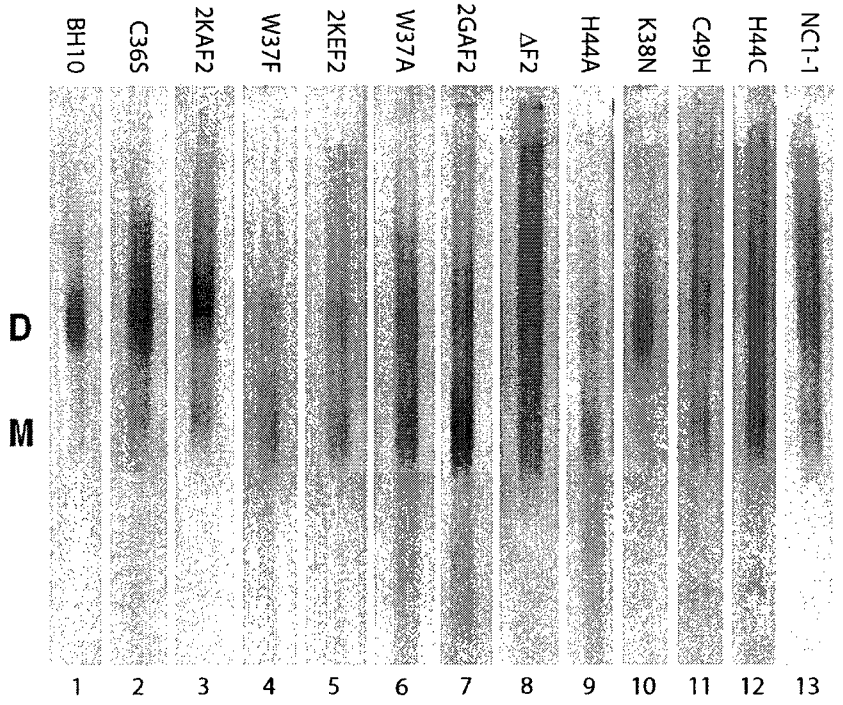


Fig. 2.5. Dimerization level of viral RNA isolated from HIV-1_{HXB2} mutated in zinc finger 2 of the nucleocapsid protein. Experimental conditions as in Fig. 1. Exposure times varied from 20 min to 16 h.

Packaging

Except for 2KAF2 and K38N, which were less effective, each mutation reduced gRNA packaging 3 to 5-fold, i.e. approximately as much as deleting the finger (Table 2). This indicates a large measure of interdependence between many residues of the finger, and a smaller importance of its highly basic residues in this respect.

Mutations ΔF2, H44C, and C49H reduced virus stability about 4-fold; 2KEF2 reduced it 7-fold. The other mutations had intermediate effects, except for W37A, which had no effect (Table 2). Overall, H44A, H44C and C49H had generally similar effects on all viral functions studied except Pr55gag processing. And 2KAF2 had milder effects than 2KEF2 on every function tested, except for comparable effects on Pr55gag processing. Little evidence of an effect of the zinc finger 2 mutations on RT packaging was seen (Table 2).

Some of the mutations had effects on gRNA packaging and virus stability that were similar to those of the corresponding mutations in zinc finger 1: 2KEF2 vs. 3EF1, K38N vs. N17K, and H44A vs. H23A, had similar pair-wise effects on gRNA packaging and virus stability; 2GAF2 and 2GAF1 had similar effects on virus stability.

Effect of linker mutations

Eight mutations were examined (Table 1). Deleting AlaPro31 (ΔAP), deleting the linker (Δlinker), lengthening it by 5 glycines (LL), neutralising it by mutating ArgLysLys34 into SerSerSer (S3) or PheValIle (FVI), or making it acidic by mutating ArgLysLys34 into GluGluGlu (S3E), had devastating effects on viral replication, while P31A and L+ (AlaPro31 replaced by LysLys) were WT-like and about 10,000 times times less infectious than WT, respectively (Fig. 2 C). S3, FVI and S3E reduced the charge of the linker from +4 to +1, +1 and -2, respectively.

Dimerization

Mutations S3, P31A, and L+ had no or little effect on gRNA dimerization (Fig. 6, lanes 3, 7, 9); S3 decreased the charge of the linker by 75% and L+ increased it by 50%. This suggests that the electric charge of the linker, as well as the bend provided to the linker by the proline (131), do not mediate gRNA dimerization. The five mutations that blocked viral replication (Fig. 2 C) impaired gRNA dimerization, and, with the exception of S3E, they did it without markedly affecting Pr55gag processing (Fig. 3C). Shortening the linker (Δlinker, ΔAP), lengthening it (LL), or making it hydrophobic (FVI), impaired gRNA dimerization to the same extent as deleting zinc fingers 1 or 2 (compare Fig. 6 lanes 2, 4, 6, 8 with Fig. 4 lane 3 and Fig. 5 lane 8; Table 2), and mutation S3E was even more damaging: it reduced gRNA dimerization to 45 % of WT (Fig. 6 lane 5; Table 2). This is

much more disabling than deleting zinc fingers 1 or 2, and more disabling than protease inactivation, which reduced dimerization to 60% of WT (Fig. 3; Table 2; (253)). Moreover, S3E had a smaller impact on Pr55gag processing than R7E and ΔF1 (Table 2; Figs. 3 A lane 4, 3B lane 11, and 3 C lane 9), two mutations that reduced gRNA dimerization to 79 % and 72.5 % of WT, respectively (Table 2). These comparisons suggest that > 50 % of the effect of S3E on gRNA dimerization was unrelated to its effect on Pr55gag processing, and that the gRNA dimerization seen in PR^- virions might be at least partly NC dependent (Jalalirad, Kafaie, Song & Laughrea, in progress). The inhibitory effect of ΔAP was not due to the absence of Ala30 or Pro31 because mutations P31A and L+ did not impair gRNA dimerization.

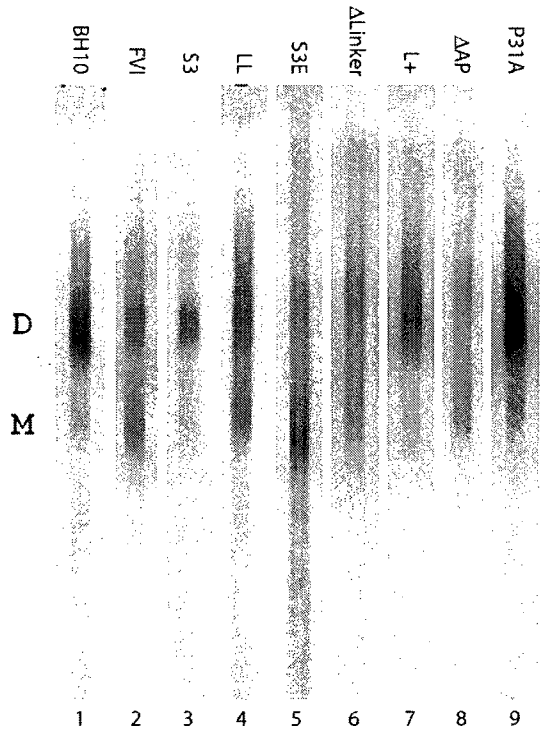


Fig. 2.6. Dimerization level of viral RNA isolated from HIV-1_{HXB2} mutated in the linker segment of the nucleocapsid protein. Experimental conditions as in Fig. 5.

Packaging

Except for S3, LL and FVI, which were less effective, the linker mutations reduced gRNA packaging 3 to 6-fold. Deleting the linker, mutating it so that it contained 6 basic residues in a row (L+), or making it negatively charged (S3E), were as effective as deleting the proximal or the distal finger (Table 2). Note the similarities between L+ and N+: they impaired only gRNA packaging among the specific viral functions studied. The wild type like viral replication of P31A (Fig. 1 C) suggests that this mutation had little effect on viral functions other than gRNA packaging.

RT packaging was reduced 10-fold by S3E; ΔAP had no effect and LL had a 2-fold inhibitory effect. Some of the effect of S3E might conceivably be due to incomplete or premature (e.g. cytoplasmic) Pr160gag-pol processing. Substituting leucine for Pro31 was previously shown to block RT packaging and incorporation of Pr160gag-pol in HIV-1_{NL4-3} (107,221). Virus stability was reduced 4 to 8- fold by Δlinker, ΔAP, and S3, depending on the mutation. LL had an intermediate 2-fold effect, while S3, P31A and L+ had no or little effect.

Joint mutations in two NC segments

Mutation NL (R7 + FVI combined) impaired the N-terminus and the linker; FL (H23A + FVI combined) impaired zinc finger 1 and the linker; and mutation FF (H23A + 2GAF2) impaired zinc fingers 1 and 2 (Table 1). In these mutants, mutations were combined that strongly inhibited gRNA dimerization (e.g. R7 but not N+; H23A but not N17K; 2GAF2 but not 2KAF2) while imposing small changes to the primary structure of NC (R7 but not R7E; H23A but not 3EF1; 2GAF2 but not 2KEF2). Mutations CH, HC and NC2-1 impaired zinc fingers 1 and 2, by combining H23C + C49H, C28H + H44C, and NC2-2 + NC1-1, respectively (Table 1). Here, mutations were combined that might preserve much of the tertiary structure of NC while having non-negligible effects on gRNA dimerization. In NC2-1, the positions of the two zinc fingers were interchanged. The results are

shown in Fig. 7 and Table 2. All joint mutations studied blocked viral replication (Fig. 2 D). Mutations R7 + FVI (NL) had additive effects on gRNA packaging but they did not have additive effects on gRNA dimerization (Table 2). This suggests that the basic residues of the N-terminus and the linker play connected roles in gRNA dimerization. Mutations H23A + FVI (FL), as well as H23A + 2GAF2 (FF), had additive effects on gRNA dimerization and on gRNA packaging (Table 2), suggesting that zinc finger 1 and the linker, as well as fingers 1 and 2, have roles in gRNA dimerization and packaging that can be at least to some extent dissected apart: FVI would not affect the dimerization and packaging functions disabled by H23A, while H23A would not affect the functions impaired by 2GAF2. Mutants FL and FF reduced gRNA dimerization more effectively than protease inactivation, but only the FL result was significantly lower than PR$^-$ (Table 2; $P < 0.05$).

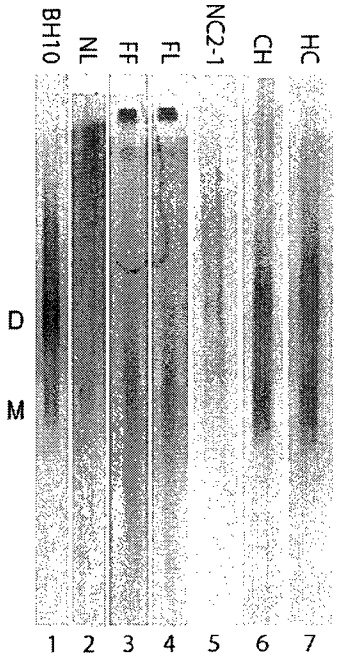


Fig. 2.7. Dimerization level of viral RNA isolated from HIV-1_{HXB2} mutated in two separate segments of the nucleocapsid protein. Experimental conditions as in Fig. 1. Exposure times varied from 15 min to 16 h.

Mutations C28H + H44C (HC) had additive effects on gRNA dimerization, but their effects on gRNA packaging were consistent with both additivity and non-additivity (Table 2). It could not be resolved whether the effects of H23C + C49H (CH) on gRNA dimerization and packaging were additive or not (Table 2). The effects of finger interchange were not additive. In mutant NC2-1, gRNA packaging was comparable to WT or NC1-1, and gRNA dimerization was similar to what was seen in NC1-1 or NC2-2 (Table 2; Fig. 7). Overall, the zinc finger deletion and interchange experiments indicate that: 1) the absence of one zinc finger is as disabling as the absence of the other (compare ΔF1 with ΔF2); 2) for gRNA dimerization, 2 fingers are better than one no matter their identity or order (compare NC2-1, NC1-1 and NC2-2 with ΔF1 and ΔF2); 3) for gRNA packaging, 2 fingers are better than one as long as one of the two is zinc finger 1 (compare NC2-1 and NC1-1 with ΔF1, ΔF2 and NC2-2; NC2-2 packaged gRNA non significantly better than ΔF1 or ΔF2 [$P > 0.1$]).

Virus stability and RT packaging were each reduced about 4-fold in mutants NL, CH and HC, and virus stability was reduced 3 to 4-fold in mutant FL. Like NC1-1 and NC2-2, NC2-1 had about a 2-fold inhibitory effect on virus stability and RT packaging. Mutations H23C + C49H (CH) and C28H + H44C (HC) were synergistic regarding RT packaging (Table 2), even though Pr55gg processing did not seem further impaired by the double mutations (compare Fig. 3E with Fig. 3B and D).

2.4. Discussion

Overview

Our main results can be briefly summarized as follows. 1) The amino-terminus, the proximal zinc finger, the linker, and the distal zinc finger of NC each contribute roughly equally to efficient

gRNA dimerization. 2) The N-terminal and linker segments appear to play predominantly electrostatic and steric roles, respectively. 3) Mutating the hydrophobic patch or the conserved glycines of the proximal or the distal zinc finger is as disabling as deleting the corresponding finger, while mutating the CCHC motif of either finger, interchanging the zinc fingers, or replacing one zinc finger by a copy of the other one, has generally intermediate effects. 4) An acidic linker, or a joint mutation in the proximal finger and the linker, can impair gRNA dimerization more than an inactive viral protease. 5) The charge of NC can be increased or decreased by up to 18% without affecting gRNA dimerization. 6) Each segment of NC plays a role, direct or not, in gRNA dimerization, gRNA packaging, and virus stability. 7) Sixteen mutations inhibited gRNA packaging more than 3-fold; eleven mutations inhibited gRNA packaging less than 3-fold, yet blocked viral replication. 8) Some mutations, involving no more than two or three NC residues, reduced virus stability 5 to 6-fold; this is not far from the 10-fold reduction previously seen as a result of jointly mutating 10 highly basic NC residues (279,280).

Regarding gRNA dimerization, the effects of 38 and 40 of the NC mutations were studied for the first time in HIV-1 and in HIV-1_{HXB2}, respectively. Regarding virus stability and RT packaging, the effects of all mutations were novel. Regarding gRNA packaging, the effects of 20 and 32 of the mutations were studied for the first time in HIV-1 and in HIV-1_{HXB2}, respectively. The effect on packaging of 8 mutations, though previously studied in HIV-1_{HXB2} (Table 3), was examined for comparative purposes. The effect on gRNA packaging of 13 of the mutations had previously been studied in HIV-1_{NL4-3} (Table 3). Knowing the effect of these mutations in both HIV-1_{NL4-3} and HIV-1_{HXB2} can clarify the influence of genetic context on phenotype. Each NC segment appeared similarly important for gRNA packaging.

NC is implicated in late infection events (e.g. gRNA dimerization, gRNA packaging, Pr55gag processing and virus stability, all studied here) and in early infection events such as reverse transcription of gRNA and proviral DNA integration into the host genome (e.g. (138,172). Some NC mutations might accordingly have no effect on late infection events, yet impair or block viral replication. This was not seen here. Each of the 40 mutations impaired one or several of the late functions studied. However, mutations P31A and N+ had mild effects on late infection events and little impact on viral replication: they are thus unlikely to inhibit early infection events in HIV-1_{HXB2}. With the exception of borderline 2KAF2, the 38 other mutations probably inhibit both early and late infection events because their impact on viral replication was inordinately large relative to their effect on late infection events. For example, mutation NC2-1 blocked viral replication and had relatively mild effects on late infection events (Table 2), suggesting that it impairs early infection events, which it does (287,288).

Breaking the protease inactivation "barrier".

Mutations FF and HC impaired gRNA dimerization as strongly as viral protease inactivation (PR^-), while S3E and FL were significantly more disabling. S3E and FL are the only retroviral NC mutations known to inhibit gRNA dimerization more strongly than protease inactivation. This suggests that the gRNA dimerization seen in PR^- virions is stimulated by unprocessed viral proteins. For example, the NC component of Pr55gag may stimulate viral RNA dimerization. Alternatively, S3E and FL may obstruct a gRNA dimerization site in cis (at the RNA level), or in trans, e.g. via aberrant binding of mature mutant NC to gRNA. The cis option is unlikely because the NC-coding sequence can be deleted (as long as the missing proteins are provided in trans) without reducing gRNA dimerization or making the dimers thermolabile (223); other experiments indicate that,

though there may be gRNA interactions sites downstream of the 5' untranslated region (153), the NC region of gRNA is not one of them (289).

None of the mutants reduced gRNA dimerization close to zero. It cannot be ruled out that up to 40 % of the dimeric RNAs seen in the isolated virions may have accumulated spontaneously (i.e. in a protein-independent manner). Inactivating the DIS or the viral protease impairs gRNA dimerization at least two and five times more, respectively, in newly released virions than in grown-up, i.e. predominantly ≥ 10 h old, virions (253). It is possible that mutations in NC may also prove more impairing in freshly released viruses.

Losing or gaining one or two electric charges has no impact on gRNA dimerization but a generally large impact on gRNA packaging

Genomic RNA dimerization was unaffected by altering the charge of NC by up to 18 % (removing or adding one or two highly basic residues); removing three positive charges was generally disabling (R7 and 3AF1, but not S3); removing ≥ four was always disabling. Out of seven mutations that had no effect on gRNA dimerization, six involved a change in electric charge: mutations N+ and L+ added two positive charges each to NC; 2KAF2 and S3 subtracted two and three charges, respectively, while N17K and K38N added and subtracted one charge, respectively (the seventh mutation was P31A). The implicated residues (Asn5 + Gln9, Asn17, AlaPro31, ArgLysLys34, Lys38, and Lys41 + Lys47) represent 20% of NC.

A more basic NC had disparate residue-dependent effects on gRNA packaging. Mutations N+ and L+ reduced gRNA packaging 4 to 5-fold, while N17K increased gRNA packaging 2 to 3-fold (Table 2). Subtracting two to three, or four to six, positive charges from NC reduced gRNA packaging an average of 2-fold (see R7, 3AF1, S3, FVI and 2KAF2) and 6.5-fold (see R7E, 3EF1,

S3E, 2KEF2, and NL), respectively. Subtracting one positive charge was harmless in the single example studied (K38N). The three mutations that preserved or enhanced gRNA packaging, namely N17K, K38N and NC2-1, made one zinc finger more similar in sequence to the other. Mutation N+ improved virus stability.

Proline-independent steric role for the linker in gRNA dimerization

Regarding gRNA dimerization, the role of the linker appears more steric than electrostatic: a neutral and polar linker (S3) was innocuous, but a shortened (ΔAP, Δlinker), lengthened (LL), hydrophobic (FVI) or negatively charged (S3E) linker was at least as disabling as deleting zinc fingers 1 or 2. The five residues at the centre of the linker are dispensable because AlaPro31, Pro31 and ArgLysLys34 could be replaced without inhibiting gRNA dimerization. Shortening or lengthening the distance between the two zinc fingers (ΔAP, Δlinker, or LL), or disturbing their interaction via a hydrophobic linker (FVI), was half as disabling as inactivating both zinc fingers (mutation FF [Table 2]). Mutating zinc finger 1 and the linker had additive effects identical to those of mutating both zinc fingers (compare FL with FF), as if the role of the linker was to properly orient the distal zinc finger. A spatial proximity of the two zinc fingers, notably Phe16 and Trp37, has been observed (131,133); this spatial proximity influences the binding constant for zinc of each finger (282). Substituting Leu for Pro31 did not inhibit HIV-1_{NL4-3} gRNA dimerization (252).

Regarding gRNA packaging, a long linker (LL) was two to four times less disabling than a short linker (ΔAP, Δlinker). This suggests that specific finger-finger interactions are less crucial for gRNA packaging than for gRNA dimerization.

An electrostatic role for the N-terminus in gRNA dimerization

The role of the N-terminus appears to be more electrostatic than that of the linker, since reducing the charge of the N-terminus from +5 to +2 (R7) impaired gRNA dimerization, contrary to the null impact of reducing the charge of the linker from +4 to +1 (S3). Since mutations in the N-terminus and the linker did not have additive effects (compare NL with R7 and FVI), they may inactivate a common effector. In Moloney murine leukemia virus (MoMuLV), the replacement of three arginines by leucine and serines in the N-terminus of NC did not impair gRNA dimerization but strongly inhibited gRNA packaging (290).

Powerful glycine pairs

In human cellular nucleic acid binding proteins and most lentiviral NCs, $GlyX_2Gly$ immediately precedes the histidine of each zinc finger. Consistent with this high degree of conservation, we found that substituting alanines for the glycines, in any one of the zinc fingers, inhibited gRNA dimerization and packaging as strongly as deleting the finger (Table 2). There is extensive hydrogen bonding within the N-terminal (125) and the C-terminal (291) zinc fingers. The presence of sterically nondemanding glycines at these positions might be essential for a stable finger structure. In Rous sarcoma virus, the distal glycine of the N-terminal zinc finger could be replaced by alanine without affecting gRNA packaging and dimerization; replacement by a bulkier valine reduced packaging by 90% without affecting dimerization (255). In simian immunodeficiency virus strain mne (SIVmne; isolated from Macaca nemestrina), the distal glycine of zinc fingers 1 or 2 could be replaced by aspartic acid without affecting gRNA packaging (171). In MoMuLV, replacement of the same glycine by valine strongly inhibited gRNA packaging (292).

Powerful aromatic residues

In each zinc finger of lentiviral NCs and human cellular nucleic acid binding proteins, an aromatic residue adjoins the first cysteine (269,293,294). When NCp7 interacts with the SL2 or SL3

stem-loops of gRNA, Phe16 and Trp37 form hydrogen bonds with guanosines from each apical loop (106,135). The point mutations F16A, W37A and W37F inhibited gRNA dimerization and packaging as strongly as deleting the fingers to which they belonged (Table 2).

In MoMuLV NC, replacing an aromatic residue of the zinc finger by glycine, serine or leucine strongly reduced gRNA packaging (119,292) without impairing gRNA dimerization (292). However, protease inactivation does not reduce the percentage of gRNA dimerization in MoMuLV (256,295-297) but strongly reduces it in HIV-1, be it HIV-1_{HXB2} produced by Cos-7 (252), HeLa ((253,256); Table 2), or 293T cells (253,298,299), or HIV-1_{NL4-3} (286,295). Thus mutating the aromatic residues of the zinc fingers of HIV-1 and MoMuLV NC disabled gRNA dimerization comparably, i.e. no more than inactivating the protease (Table 2). Since very few NC mutations can impair gRNA dimerization more strongly than protease inactivation (Table 2 and (252)), it is not surprising that none of 10 MoMuLV NC mutations reduced the gRNA dimerization yield (256,290,292,300).

Zinc finger deletion

Deleting the N-terminal or the C-terminal zinc finger reduced gRNA dimerization similarly (Table 2). Analogous results were obtained in Rous sarcoma virus (Méric *et al.*, 1988). This is consistent with experiments indicating that the proximal and the distal zinc fingers of NC each play an important role in the in vitro stabilization of a short murine retroviral RNA dimer (91).

Deleting zinc finger 1 reduced gRNA packaging by dissimilar amounts in HIV-1_{HXB2}, HIV-1_{NL4-3}, and SIVmne (Table 3). Deleting zinc finger 2 reduced gRNA packaging by approximately 75% in HIV-1_{HXB2} and 35 to 55 % in SIVmne (Table 3). Interstrain differences were also seen when a functional homologue of finger deletion, i.e. the replacement of 2 or 3 cysteines of one finger by serines, was studied. When in finger 1, such mutations reduced gRNA packaging by 90 to 98 % in

HIV-1_{NL4-3} (277,301,302), and by 5, 74, or 95% in various SIV strains (303-305). When in finger 2, they reduced gRNA packaging by 0 (301), 27 (302), or 80 % (277) in HIV-1_{NL4-3}, and by 20 to 97 % in various SIV strains (303-305). When the first two cysteines of both zinc fingers were replaced by tyrosines, HIV-1_{HXB2} gRNA packaging was reduced by 75-80 % (51) to > 99 % (274), and HIV-1_{NL4-3} gRNA packaging was reduced by 93 % (302). The origin of these variabilities is unknown but may be partly related to variant quantities of proviral DNA per transfected cell.

Zinc finger interchange or replacement of one finger by a copy of the other

These three mutations inhibited gRNA dimerization approximately 60% as well as zinc finger deletion (Table 2). Interchanging the zinc fingers or replacing finger 1 by a copy of finger 2 was studied in Rous sarcoma virus, with an effect comparable to that of deleting a zinc finger (254).

Interchanging the zinc fingers or replacing finger 2 by a copy of finger 1 had no or little impact on gRNA packaging in HIV-1_{HXB2}, i.e. generally less than in HIV-1_{NL4-3} or Rous sarcoma virus (Table 3). Replacing zinc finger 1 by a copy of finger 2 reduced gRNA packaging by 30 to 85 % in HIV-1_{HXB2}, HIV-1_{NL4-3} and Rous sarcoma virus (Table 3). The qualitative retroviral consensus is: replacing zinc finger 1 by a copy of finger 2 inhibits gRNA packaging; interchanging the fingers or replacing a finger by a copy of the other inhibits gRNA dimerization.

Dysfunctional alternative zinc coordination motifs

Retroviral type zinc fingers that have been mutated into a CCCC or a CCHH motif are suspected of retaining the ability to coordinate zinc (267,275,306-308). We have shown that a CCHH motif in any of the two zinc fingers, or a CCCC motif in the distal finger, inhibited gRNA dimerization and packaging ≥ 70 % as well as deleting the corresponding finger (Table 2). While mutation H23C inhibited gRNA packaging as strongly as deleting the N-terminal zinc finger, it had

modest inhibitory effects on gRNA dimerization (~ 30 % of the effect of deleting the finger). Thus lentiviral gRNA dimerization requires a zinc coordination motif that clearly cannot be replaced by CCHH or CCCC at the distal position and by CCHH at the proximal position. One possible interpretation is that the affinity of the CCHH and CCCC motifs for zinc is generally insufficient, particularly in zinc finger 2. A CCHH motif in the sole zinc finger of MoMuLV NC did not impair gRNA dimerization and packaging (300); a CCCC motif did not impair gRNA packaging (174).

Point mutations H23A and H44A reduced gRNA dimerization as much as F16A, W37A, or deleting the fingers to which they belonged (Table 2). Interestingly, H44A causes an improper orientation of the residues forming the hydrophobic cleft that is critical for interaction of NCp7 with nucleic acids (309). For example, in NCp7 complexed with d(ACGCC), Trp37 inserts between C2 and G3 and stacks on the latter, while Phe16 stacks on the C2 ring (310). Thus, H44A and W37A have similar effects on gRNA dimerization and packaging plausibly because both displace Trp37.

Mutations C28S and C36S inhibited gRNA dimerization less efficiently than H23A and H44A, but reduced gRNA packaging as strongly as these mutations (Table 2). Though C28S and C36S considerably reduce the zinc-binding ability of the N-terminal (282) and the C-terminal (134) zinc fingers, respectively, they may in some respect be less sterically damaging than H23A and H44A, because cysteine and serine only differ by the substitution of one sulfur for one oxygen atom. Overall, the effect of C28S and C36S on gRNA dimerization and packaging was surprisingly comparable to the effect of transforming the CCHC motif into CCHH or CCCC.

Context-dependent phenotype of some mutations

We have compared the effect of 10 mutations on gRNA dimerization in HXB2 and NL4-3 virions produced by HeLa cells. We have found that mutations FVI, C49H and HC were more disabling in HIV-1_{HXB2} than in HIV-1_{NL4-3}, and that H23C, C28H, H44C, CH, NC1-1, NC2-2 and NC2-1 had comparable effects in the two isolates (data not shown). Except for FVI, these NL4-3 proviral clones were gifts from Robert Gorelick. R7 was more disabling in HIV-1_{HXB2} produced by HeLa cells (Table 2) than in HIV-1_{NL4-3} produced by Cos-7 cells (252). Overall, the results indicate that HIV-1_{HXB2} may be more sensitive than HIV-1_{NL4-3} regarding the effect on gRNA dimerization of 25% of the mutations investigated. It is also noteworthy that HIV-1_{HXB2} gRNAs were 8 ± 2% less dimeric than HIV-1_{NL4-3} gRNAs (data not shown).

HIV-1_{HXB2} gRNA packaging seemed more impaired by some mutations and less by others. Mutations H44C, C49H, and CH were more disabling in HXB2 than in HIV-1_{NL4-3}; ΔF1, S3, HC, and NC2-1 were less disabling in HIV-1_{HXB2} than in HIV-1_{NL4-3}; and R7, C28H, C36S, NC2-2 and NC1-1 had comparable effects in both isolates (Table 3). Substituting AlaAla for ArgLys11 was more disabling in HIV-1_{HXB2} than in HIV-1_{NL4-3}. (271) (H23C could not be compared because of discordant effects yielding NL4-3 results on either side of the HXB2 data [Table 3]). Thus 5 mutations out of 13 showed no evidence of context-dependent gRNA packaging; the context-dependence of the remaining mutations favored neither HXB2 nor NL4-3 in terms of phenotypic robustness. We conclude that 25 to 50% of NC mutations may display a context dependent gRNA dimerization or packaging yield; gRNA dimerization, but not gRNA packaging, appears to be somewhat more labile in HIV-1_{HXB2} than in HIV-1_{NL4-3}.

2.5. Materials and Methods

Plasmid construction. Proviral vector pSVC21.BH10 encodes a HIV-1_{HXB2} cDNA clone. Mutant proviral vectors were constructed from pSVC21.BH10 by PCR mutagenesis, using primers

described in Table 4. The nucleotides positions are based on the sequence of HIV-1 gRNA. To prepare mutants in the N-terminus and zinc finger 1 regions, a DNA fragment extending from Spe I to Apa I restriction sites was synthesized with desired mutations by PCR, and ligated into pSVC21.BH10. Mutants in the linker and zinc finger 2 regions were constructed by PCR from a DNA fragment extending from Spe I to Bcl I, and from Apa I to BclI, respectively. FF, NL and FL were constructed using mutants H23A, 2GAF2, R7 and FVI. They were cut using restriction enzymes Apa1 and Spe1 (Amersham). The H23A DNA fragment was inserted into 2GAF2 to construct FF, and in FVI to prepare FL. The same DNA sequence from R7 was introduced in FVI to make NL. After mutagenesis and ligation, all mutated DNA fragments produced by PCR were completely sequenced (ACGT Inc., Toronto) to verify that the wanted mutation, and no other mutation, was introduced by the mutagenic procedure.

Cell culture and transfections. HeLa Cells were cultured at 37 °C in a medium consisting of Dulbecco's modified Eagle's medium (DMEM), 10 % fetal calf serum, ampicillin and streptomycin (Invitrogen). The PolyFect transfection reagent (Qiagen) was used to transfect 9 μg of proviral DNA into 50 % to 70% confluent HeLa cells in 100- by 20-mm petri dishes containing 10 ml of culture medium.

Infectivity assay. Mutant proviruses and the parental BH10 provirus were independently transfected into HeLa cells. Virus-containing supernatants were collected 48 hrs post transfection and passed through 0.2 μm pore-size cellulose acetate filters to remove the cells. The CAp24 content of these clarified supernatants was measured using an ELISA kit (Vironostika HIV-1 Antigen, Biomerieux). Equal amount of the supernatants (10 ng of CAp24 content) were used to infect equal numbers of MT2 cells (6 x 10^6 cells in 10 ml of RPMI 1640 medium, 10 % fetal calf serum, ampicillin and streptomycin (Invitrogen), per petri dish). In the human T-cell line MT2, only a short

time lag separates infection from viral replication. After 2 h, cells were washed twice to remove unbound viruses and were then maintained in serum-supplemented medium. On every other day, cells were diluted 1 in 2 into fresh medium and the RT activity in the supernatant of the removed medium was determined. RT activity measurements were made over a period of 14 days.

RT activity. The exogenous (oligo (dT) directed) RT activity was measured by adding 40 µl of RT cocktail (60 mM Tris-HCl [pH 7.9], 180 mM KCl, 6 mM $MgCl_2$, 6 mM dithiothreitol, 0.6 mM EGTA, 0.12 % Triton X-100, 6 µg/ml oligo (dT), 12 µg/ml poly(rA), 0.05 mM ^{3}H dTTP) to a 10 µl sample. After incubation for 2 h at 37 °C, the reaction was stopped with cold 10% TCA (150 µl per well), and precipitated for 30 min at 4 °C. The precipitate was blotted, washed and scintillation counted.

Virus purification and isolation of HIV-1 viral RNA. Filtered virus-containing supernatants were centrifuged (SW41 rotor, 35 000 rpm, 4 °C, 1 h), through a 2 ml 20 % (w/v) sucrose cushion in phosphate-buffered saline (PBS). The virus pellet was dissolved in 400µl sterile lysis buffer [50 mM Tris (pH7.4), 50 mM NaCl, 10 mM EDTA, 1 % (w/v) SDS, 50 µg tRNA per ml, and 100 µg proteinase K per ml], and extracted twice at 4 °C with an equal volume of buffer-saturated phenol-chloroform-isoamylalcohol (25:24:1) (Invitrogen). The aqueous phase was precipitated overnight at -80 °C with 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of 95 % ethanol, and centrifuged at 14,000 rpm in an Eppendorf 5145 micro centrifuge at 4 °C for 30 min. The gRNA pellet was rinsed with 70 % ethanol, and dissolved in 10 µl buffer S (10 mM Tris (pH 7.5), 100 mM NaCl, 10 mM EDTA and 1 % SDS) (253).

Electrophoretic analysis of HIV-1 gRNA. The gRNA was electrophoresed under non denaturing conditions and identified by Northern (RNA) blot analysis (253). Electrophoretic conditions were 4 V/cm for 4 h on a 1 % (w/v) agarose gel in TBE2 (89 mM Tris, 89 mM Borate

and 2 mM EDTA, pH 8.3) at 4 °C. After electrophoresis, the gel was heated at 65 °C for 30 min in 10% (w/v) formamide, and the embedded RNAs were diffusion transferred to a Hybond N+ nylon membrane (Amersham). After drying at room temperature for 2 h, crosslinking (3000 j in a UV Stratalinker), and prehybridization at 42 °C for 3 h in 6X SSPE (1X SSPE is 0.15 M NaCl, 10 mM NaH2PO4, and 1 mM EDTA, pH 7.4), 50% (w/v) deionized formamide, 10% dextran sulfate, 1.5% SDS, 5X Denhardt's reagent, 100 μg/ml salmon sperm DNA, the membrane was hybridized overnight in prehybridization buffer devoid of Denhardt's reagent in a rotating hybridization oven at 42 °C to approximately 25 μCi of ^{35}S-labeled antisense RNA 636-296 (a 356-nt RNA that is the antisense of the 296 to 636 region of the HIV-1 genome prepared with the SP6 Megascript kit [Ambion]) (311). This was followed by two 30 min washes in 1X SSC [1X SSC is 0.15 M NaCl plus 0.015 M sodium citrate]- 0.1% SDS at room temperature and 37 °C, and one 30 min wash in 0.2X SSC- 0.1 % SDS at 45 °C (246), exposure to a Kodak BioMax MR X-ray film, and densitometric analysis.

Densitometric analysis. The autoradiograms were scanned and analysed with the NIH 1.6.3 program. Care was taken to scan variously exposed films to guard against over-exposed or under-exposed bands or spots. The monomer and dimer bands were considered of equal width; that width was approximately twice the vertical size of the D and M letters used to indicate dimers and monomers in the relevant figures. Material located elsewhere in the gels was not taken into account in the calculation of the percentage of dimers. The diffuse character of many bands may reflect conformational diversity among the gRNA molecules. It is not due to poor resolution of the gels because heat denatured gRNAs formed a sharp band at the monomer position (not shown). The RNA bands seen here are not more diffuse, and sometimes sharper, than what was seen in earlier studies by us and others (223).

Genomic RNA packaging. The amount of gRNA per unit CAp24 of virus was quantitated by hybridization with antisense RNA 636-296 using a dot blot assay. Virus pellets were resuspended in 400 µl of Trizol LS reagent (Invitrogen), and incubated at 30 °C for 5 min. 100 µl of chloroform was added, followed by shaking for 15 seconds and incubation at room temperature for 15 min. After centrifugation (12,000 × g, 15 min, 4 °C), the colorless aqueous upper phase was mixed with 250 µl of isopropyl alcohol, incubated at room temperature for 10 min and centrifuged again. The precipitated RNA was washed once with 500 µl of 70% ethanol, pelleted (7500 × g, 5 min, 4 °C) , air-dried, dissolved in 10 µl RNase-free water and stored at -20 °C. Serial 10-fold dilutions of wild type RNA samples, normalized for input virion CAp24, were used to construct a standard curve. 29 µl of buffer F (100 % deionized formamide, 20 µl; 20x SSC, 2µl; 37 % formaldehyde, 7 µl) was added to each sample, followed by incubation (68 °C, 15 min) and chilling on ice. After adding 78 µl of 20x SSC buffer, samples were vacuum-suction transferred to a Hybond N^+ nylon membrane (Amersham) sandwiched within a Hybri-Dot filtration manifold (Bethesda Research Laboratories). The wells were washed twice with 1 ml of 10x SSC, and suction continued for a further 5 min to dry the membrane. The membrane was removed, dried for 4 h, cross-linked, pre-hybridized, hybridized, autoradiographed and scanned as for Northern blot analysis (above). To confirm the scans, each individual spot of the nylon membrane was excised and scintillation counted.

Virus stability and RT packaging. 1.2 ml of filtered virus-containing supernatant was pelleted through a 0.3 ml 20% sucrose cushion in the TL-100 Beckman ultracentrifuge (TLA 55 rotor, 45 000 rpm, 1 h, 4 °C). The virus pellet was dissolved in 10 µl of PBS and its CAp24 content was measured using an ELISA kit (Vironostika HIV-1 Antigen, Biomerieux). The CA content of the purified viruses divided by the CA content of the 48 h culture supernatant, relative to the ratio found

in WT, was taken as a measure of virus stability (279,280). The ratio pellet/supernatant was 0.43 ± 0.07 in WT (this was taken to mean 100 in Table 2).

The exogenous RT activity of the pelleted viruses divided by their CA content, relative to the ratio found in WT, was interpreted as RT packaging.

Pr55gag proteolytic maturation. At 48 hours post-transfection, cells were lysed in ice-cold NP-40 containing buffer (100 mM NaCl, 10 mM Tris, 1 mM EDTA, 0.5% NP-40, and protease inhibitor cocktail [Roche]). Supernatants were cleared by centrifugation at 3 000 g x g and filtered (0.22μm). Viruses were then concentrated through a sucrose cushion by ultracentrifugation. Equal amounts of viruses (judged by CAp24-ELISA) were lysed in the above buffer and subjected to SDS polyacrylamide gel electrophoresis. Viral proteins were detected by Western blotting using a rabbit anti-capsid antibody (ABT-Trinity Biotechnology, CA, USA) that recognizes Pr55gag, CAp24, and CA-containing partially cleaved Gag proteins, and an enhanced chemiluminescence Western blot detection kit (Amersham). The signals for CA-containing proteins were quantitated by densitometric scanning and analysed with the NIH 1.6.3 program. The signal obtained from the CAp24 band was divided by the total signal obtained from all CA-containing proteins, to calculate Pr55gag processing level.

Chapter 3

Nucleocapsid Maturation and Number Requirements For HIV-1 Genomic RNA Dimerization

Jafar Kafaie, Marjan Dolatshahi, Lara Ajamian, Rujun Song, Andrew J. Mouland, Isabelle Rouiller, Michael Laughrea. Role of capsid sequence and immature nucleocapsid proteins p9 and p15 in Human Immunodeficiency Virus type 1 genomic RNA dimerization. Virology. 2009 Mar 1; 385 (1): 233–244. Epub 2008 Dec 13.

This manuscript contains Cryo-electron-microscopy studies that are not included in this thesis.

Preface to Chapter 3

NC has a role in viral gRNA dimerization, infectivity, RT packaging, viral stability, and gRNA packaging as I explained in the second chapter to my thesis and other references therein. NC is first translated as part of a larger precursor Pr55gag polyprotein which then gets processed to smaller proteins containing NC, namely p15, p9 and p7. The role of these intermediate proteins during the viral life cycle remains largely unknown. The minimal amount of protease enzyme required for wild type-like gRNA dimerization and the quantitative relationship between Pr55gag processing and gRNA dimerization was unknown in retroviruses. In the second part of my project we introduced point mutations in scissile bonds of the Pr55gag polyprotein in order to eliminate cleavage at the p7-p1 and/or the p1-p6 junctions, and produce viruses with intermediate NC proteins and study their effect on HIV-1 gRNA dimerization. Finally, we designed co-transfection assays to produce composite viruses with altered protease activity and we studied the relationship of Pr55gag processing and gRNA dimerization.

3.1. Abstract

During viral assembly, the Pr55gag polyprotein of human immunodeficiency virus type 1 (HIV-1) is proteolytically processed into matrix (MA), capsid (CA), SP1/p2, nucleocapsid (NC), SP2/p1 and p6 proteins by the viral protease. The first step in viral protein processing starts with the release of the viral protease from Gag-Pol polyprotein and the broad range of cleavage efficiencies lead to stepwise and ordered processing of the Pr55gag polyprotein. Newly released viral particles lack dimeric genomic RNA (gRNA), which then increases after release of the virus. Dimerization of gRNA seems to be important for viral infectivity and *in vivo* experiments have shown that mutations in N-terminus, Linker or zinc fingers 1 and 2 of NC protein can influence gRNA dimerization and packaging, infectivity and stability. In this paper we first studied the effect of Pr55gag processing on gRNA dimerization by introducing point mutations that minimally modify the Pr55gag p7-p1 and p1-p6 scissile sites in order to block or enhance cleavage. Western blot studies proved the *in vivo* alteration of Pr55gag processing and the formation of viral particles containing no NCp7 (in mutant p9) or no NCp7 and no NCp9 (in mutant p15) nucleocapsid protein. While mutations p9 and p15 both yielded noninfectious particles, mutation p9 had almost no effect on gRNA dimerization while mutation p15 reduced it to 75% of wild type. The minimal requirement of protease activity needed to produce a wild type level of gRNA dimerization was studied by site directed mutagenesis of the protease and also by changing the proportion of BH10 and PR^- (protease inactivated) proviruses in cotransfection studies. We found that only a 25 % processing of Pr55gag is needed in 48 hours old viruses to produce a wild type level of gRNA dimerization. We also used equimolar cotransfection studies by BH10 and NC mutated proviruses and found that 50% wild type NCp7 in nucleocapsid vicinity is enough to produce wild type level gRNA dimerization in NC defected mutants.

3.2. Introduction

The Pr55gag and the Gag-pol polyproteins of HIV-1 get rapidly processed into an N-terminal component (the matrix (MA)-capsid (CA)-p2 polyprotein) and two C-terminal components. The nucleocapsid (NC) NCp7-p1-p6 polyprotein is produced from Pr55gag and the Gag-pol polyproteins gives rise to a protein that starts with the NCp7 sequence and can be up to 120kDa in size if it includes the protease (PR)-reverse transcriptase (RT)-integrase (IN) sequence (312-314). This paper will focus mainly on the NCp7-p1-p6 polyprotein (NCp15), and its partial maturation products which represent 95 % of the Gag and Gag-pol C-terminal components generated. *In vitro* reaction of purified HIV-1 protease with Pr55gag-containing rabbit reticulocyte lysate (i.e. in the presence of large quantities of RNA) indicates that Pr55gag is first processed into MA-CA-p2 and NCp15. It also shows that NCp15 and MA-CA-p2 are simultaneously processed into NCp7-p1, p6, MA and CA-p2 in a stage 2 cleavage process and that NCp7 and p1 appear at about the same time as CAp24 and p2, in a stage 3 cleavage process (Fig. 1) (50,140,315).

Interest in NCp15 and its proteolytic products is generated in part from the study of HIV-1 mutants resistant to viral protease inhibitors (PI). Resistance is generated when mutations in protease reduce the binding of PIs to this enzyme. These mutations often result in impaired protease activity even in the absence of inhibitor (316-318). This is often partially relieved by suppressor mutations at the p7-p1 and/or p1-p6 junctions (319,320) that are associated with more scissile junctions (321,322) and faster replicating PI-resistant virions (319-322). Since processing at the p7-p1 and p1-p6 junctions is slow (see below), and plausibly rate-limiting after the protease has become PI-resistant, it can be readily enhanced by appropriately mutating them to more scissile junctions. This apparent necessity for a timely maturation of NCp15 drew our attention to the respective roles, or deficiencies, of NCp15, NCp9 and NCp7 in virus function, with special emphasis on their role, or deficiency, in gRNA dimerization.

The timing of the NCp15 maturation process in the isolated virus is not well known. Studies have shown that this process happens early during the viral life cycle (313,314,323,324), and that NCp15 is fully processed in 3 to 4 day old viruses (325). There are few *in vivo* data on this topic, but they are consistent with the idea that proteolytic processing of NCp15 occurs later than the separation of MA from CA (324,326), meaning that NCp15 may be longer lived than indicated by most *in vitro* data. For example, in acutely infected CEM cells, an 80 % conversion

of Pr55gag into CA-p2/CA is seen in the cytoplasm, as well as a 50 % conversion of CA-p2 into CA, but no conversion of NCp15 into p6 is seen; that conversion occurs only in the membrane and virus fractions (326). MA-CA processing is RNA-independent, while NCp15 processing is approximately 10 times slower in the absence of RNA (327). Thus *in vitro* data also allow the possibility that NCp15 may be processed at a rate similar to that of CA-p2 (the 9X of Fig. 1, derived in the presence of RNA would then become 90X). Interestingly, the k_{cat}/K_m for protease-directed hydrolysis of peptides mimicking the p1-p6 cleavage site is ≥ 50 and ≥ 2 times lower than for mimics of the MA-CA and the CA-p2 cleavage sites, respectively (318,328). This supports the idea that NCp15 processing is more a stage 3 than a stage 2 maturation process (Fig. 1), and suggests that a study of the role of NCp15 and NCp9 in the viral replication cycle may shed light on a longer than expected maturation period of HIV-1, and on the reason behind these late processing steps.

Proteolytic processing of the p2-NC cleavage site is critical for HIV-1 RNA dimer maturation (298). We will pursue this line of inquiry by asking to which extent proteolytic processing within NCp15 is critical for HIV-1 RNA dimer maturation: notably, how far can the viral RNA dimerization process go in the absence of proteolytic processing at the p7-p1 junction (i.e. in the absence of NCp7), or at the p7-p1 and p1-p6 junctions (i.e. in the absence of NCp7, NCp9 and p6).

Co-transfection assays have shown that 20 % of functional NC proteins is sufficient to yield WT-levels of gRNA encapsidation, and that 5 % of functional NC can still yield a 50 % gRNA encapsidation level (302). To characterize the role of mature NC in HIV-1 gRNA dimerization, we will use both cotransfection assays and mutations partially inactivating the viral protease, to address the question of the minimal number of mature NC needed for full gRNA dimerization. To bring context to the results, we will also study the effect of most mutations on genomic RNA packaging, virus stability and reverse transcriptase packaging.

3.3. Results and Discussion

HeLa cells were transfected in parallel with equal amounts of pSVC21.BH10 or mutant proviral vectors. Proviral vector pSVC21.BH10 encodes an infectious HIV-1_{HXB2} molecular clone derived from the IIIB strain of HIV-1 (246). After 48 h, viruses were isolated from the

culture supernatant, their capsid protein (CA) and reverse transcriptase (RT) content was measured, and their gRNA was extracted, electrophoresed on a non-denaturing agarose gel and visualized by Northern blotting with a ^{35}S-labeled HIV-1 riboprobe, followed by autoradiography. Prior to virus purification, a small volume of culture supernatant was kept to measure its CA content and determine viral replication per unit of supernatant CA (Materials and methods). The identity of the various mutants is described in Table 1. For comparative purposes, the gRNA dimerization and Pr55gag processing seen in protease-inactive (PR$^-$) virions was also included. The aspartic acid at position 25 of the viral protease active site was replaced by arginine in PR$^-$ virions; gRNA dimerization had previously been investigated in this PR$^-$ context (30).

	Construct name[a]	Viral replication[b]	gRNA dimerization[c]	gRNA packaging[c]	RT packaging[c]	Virus stability[c]
1	HXB2	+	100	100	100	100
2	PR-	-	60 ± 3	Nd	Nd	nd
3	p9	-	97 ± 1.3	108 ± 23	110 ± 5	120 ± 10
4	p1p6	+/--	97 ± 1	133 ± 45	115 ± 5	58 ± 5
5	p15	-	76 ± 7	106 ± 14	90.5 ± 7.5	70 ± 7
6	55F	+/-	98 ± 1	75 ± 7.3	97.5 ± 2.5	57 ± 2

Table 3. 1. Effect of mutations introduced into the scissile sites of gag-p55 protein and protease activity reducing mutants of HIV-1_{HXB2} on virus infectivity, genomic RNA dimerization, genomic RNA packaging, packaging of reverse transcriptase activity, virus stability and Pr55gag proteolytic maturation.

[a] Mutants are depicted in Fig.1 and explained in the results and discussion section, except for PR$^-$, which stands for protease-inactive virions: the viral protease active site aspartic acid residue at position 25 was replaced by arginine (30).

[b] Viral replication is categorized as follows. +: identical or close to wild type; ++/- equivalent to 100 times diluted wild type; +/-: equivalent to 10 000 diluted wild type; +/-- :equivalent to wild type diluted more than 10 000 times; -: no viral replication detected.

[c] For each mutant, the error bars designate the standard error of 3 to 6 independent experiments for gRNA dimerization (4 on average), 2 to 5 independent experiments, for gRNA packaging (3.5 on average), and 2 independent experiments for RT packaging and virus stability. The values for HXB2 are arbitrarily set at 100, and the values for the mutants are expressed as % of wild-type level. Genomic RNA dimerization numbers were obtained by densitometric analysis. ND: not done, or less than 2 independent experiments.

For each mutant, viral replication, gRNA dimerization, gRNA packaging, RT packaging, and virus stability was typically measured. Genomic RNA packaging was measured by dot–blot hybridization. RT packaging was defined as CA-normalized RT activity: the RT activity of isolated virions divided by their CA content, relative to the ratio found in WT samples. Unprocessed Gag-Pol, though it contains RT, is not recorded by the RT packaging assay. Virus stability was defined as CA content of purified viruses divided by CA content of the culture supernatant, relative to the ratio found in WT samples (Materials and methods).

Maturation of NCp15 into NCp9 is essential for fast rates of genomic RNA dimerization; maturation of NCp9 into NCp7 has no incidence on genomic RNA dimerization but is essential for viral replication

To evaluate the activity of NCp7 precursors NCp9 (NCp7-p1) and NCp15 (NCp7-p1-p6), as well the effect of modulating the rate of proteolytic maturation of NCp15, mutants p9, p15, p1p6 and 55F were constructed (Fig. 1 and Materials and Methods). These mutants were produced by mutating the P1 (the amino acid residue immediately upstream of the scissile bond) positions of NCp15 (Asn55 of NCp7 and Phe16 of p1) in order to modify the rate of cleavage at the NCp7-p1 and p1-p6 sites, respectively (315). We were careful to choose mutations that interfere neither with the frameshifting site of genomic RNA nor with the structure of the downstream frameshift stimulatory stem-loop; for example, replacing Phe16 of p1 by Ile (50) changes the secondary structure of two different models of the frameshift stimulatory stem-loop (329,330).

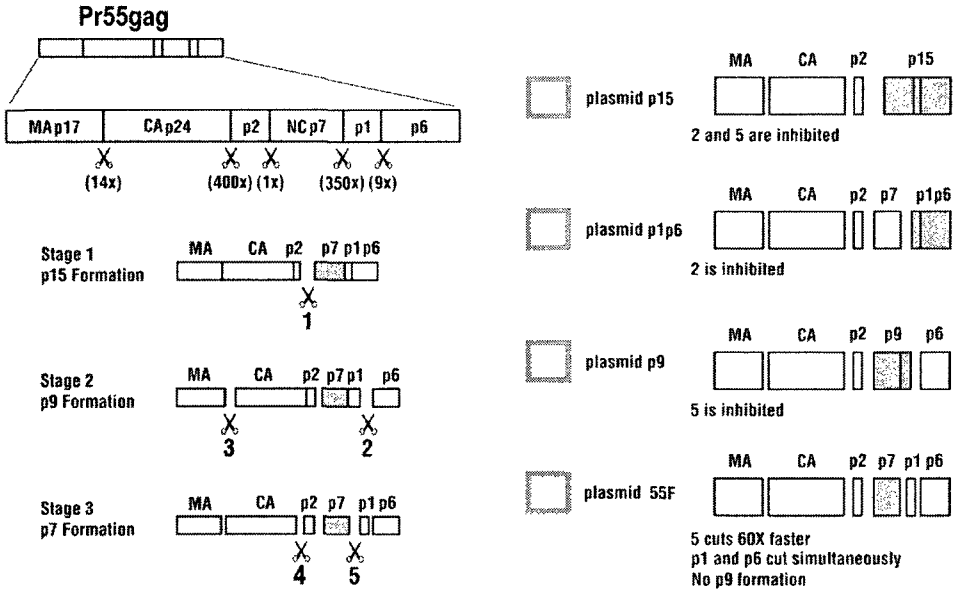


Fig. 3. 1. *In vitro* Pr55gag polyprotein processing in the presence of recombinant HIV-1 protease. Illustration of the processing steps inhibited in mutants p15, p1p6, p9 and 55F. The *in vitro* rate of cleavage indicated for 55F comes from Pettit *et al.* (315). It remains to be seen if, inside the virus, NCp15 and MA-CA-p2 are processed in parallel ways. The possibility that NCp15 is processed more slowly (later) than MA-CA-p2 has not been ruled out, and is currently under study (Jalalirad & Laughrea) Three step gag-p55 polyprotein processing with their relative cleavage speed, intermediate and final products are shown in figure 1.

The level of Pr55gag processing was assessed via Western blotting of the proteins extracted from purified mutant HIV-1, using antibodies against NCp7 (Fig. 2A) and CAp24 (Fig. 2B). Mutation p9 (Asn55 of NCp7 replaced by serine) blocked proteolytic maturation of NCp9 into NCp7 and p1, both in isolated viruses (Fig. 2A, lane 5; (315)) and *in vitro*, when recombinant HIV-1 protease reacted with the Pr55gag polyproteins (315). Mutation p1p6 (Phe16 of p1 replaced by leucine, blocking cleavage at the p1-p6 site) was designed to produce a longer-lived NCp15 intermediate (Fig. 1) that matures directly into NCp7, and at the time expected for WT apparition of NCp7 and p1. This implies delayed liberation of p6 from NCp15, stable attachment of p1 to the amino-terminus of p6, and no production of NCp9. The p1p6 mutation blocks proteolytic processing of NCp15 into NCp9 and p6 in vitro (Pettit *et al.*, 2002) and in

isolated viruses (Fig. 2A, lane 4). In addition, NCp15 proved less scissile than expected. Thus proteolytic maturation at the NCp7-p1 site was impaired either by the downstream p1p6 mutation or by the covalently linked p6 (p1 is only 16 amino acid residues long). The result was p1p6 viruses whose NC was about 75% in NCp15 form and 25% in NCp7 form, with no visible NCp9 (Fig. 2A, lane 4). In comparison, only traces of NCp9 and NCp15 were visible in BH10 (Fig. 2A, lane 2). Mutation p15 was constructed by combining the p9 and p1p6 mutations (Materials and Methods). The NC of the produced viruses was exclusively in the NCp15 form (Fig. 2A, lane 3). Mutant 55F (Asn55 of NCp7 replaced by phenylalanine) increases > 60-fold the *in vitro* rate of cleavage between NCp7 and p1 (315). Thus it should yield a more transient NCp15 intermediate, premature apparition of NCp7, and minimal amounts of NCp9 no matter the age of the virus. This earlier apparition of NCp7 was confirmed in HIV-1_{hxb2} produced by HeLa cells (315), and is consistent with the results of Fig. 2A, lane 1. In sum, Pr55gag processing was blocked at the intended cleavage sites in mutants p15 and p9. In mutant p1p6, it could not be directly verified if the inhibition was as expected, for lack of an anti-p6 antibody, but the presence of large amounts of NCp15 clearly shows that the cleavage between p1 and p6 was seriously inhibited; consistent with this is the absence of traces of NCp9 in Fig. 2 lane 5.

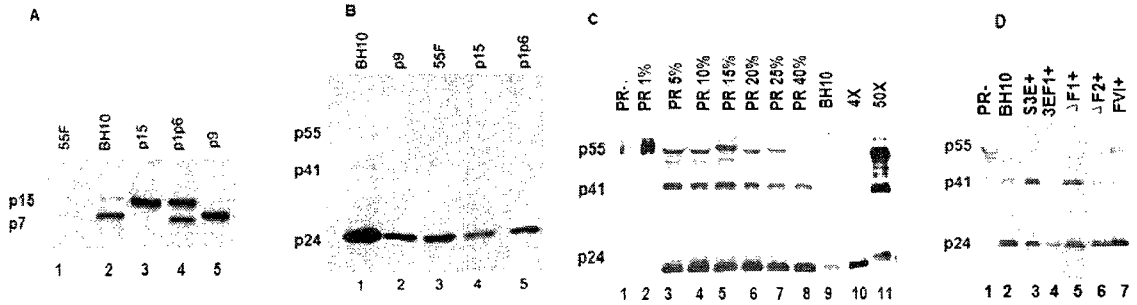


Fig. 3. 2. State of maturation of Pr55gag in HIV-1_{HXB2} mutated at the C-terminal. Gag-p55 scissile sites (A and B), produced by the cotransfection of protease-inactive and WT plasmids (C), produced by the cotransfection of NC mutated and WT plasmids (D). Proteins extracted from purified viruses were assessed by SDS-polyacrylamide gel electrophoresis, followed by visualization by Western blotting using NCp7-reactive antibodies in (A) and capsid-reactive antibodies in (B) , (C) and (D) and (Materials and Methods). With anti-capsid antibodies, Pr55gag migrates at the 55 KD position. After cleavage of Pr55gag at the primary cleavage site, and before cleavage at the secondary and tertiary sites, the capsid is part of a matrix-capsid oligoprotein running at the 41 KD position. NCp7 can not be detected by anti NCp7 antibodies in the context of Pr55gag polyprotein, but after its first cleavage is detected as p15 at 15 KD position and at 9 KD as p9. The final product of Pr55gag polyprotein processing is marked at 7 KD. Mutants are defined in Fig. 1 and in the text.

Mutations p15 and p9 blocked viral replication while 55F and p1p6 substantially delayed viral replication, i.e. by 3 to 6 days (Table 1 and Fig.3 A). Thus, without removal of p1 from NCp9 viral replication is impossible.

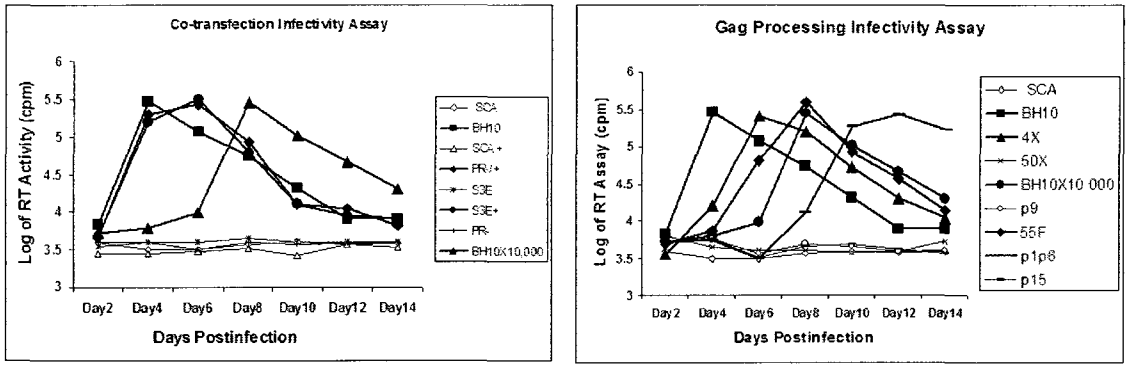


Fig. 3. 3. Viral replication of HIV-1_{HXB2} mutated in the Pr55gag processing sites, protease activity reducing mutants (4X and 50X), core eliminating mutant (A), virions derived from the co-transfections (1:1) of CA, CA+BH10, S3E, S3E+BH10, Pr- and Pr-+BH10 (B) are shown in this figure. MT2 cells were infected with an amount of undiluted progeny virus equal to 10 ng of CAp24 antigen. Virus growth was monitored by measuring reverse transcriptase activity (cpm/µl) in culture fluids at various times. The replication of 10,000X diluted wild-type HXB2 was also studied for comparative purposes. Mutants are defined in Materials and Methods.

The percentage of gRNA dimers in mutant p15 was 76% of wild type level while all other mutations had no impact on gRNA dimerization (Fig. 4 and Table 1). Thus: 1) proteolytic

maturation of Pr55gag to the NCp9 level is sufficient to stimulate stable gRNA dimerization in spite of the block in viral replication (p9 results); 2) full gRNA dimerization can occur despite a 75% fewer NCp9 and NCp7 relative to WT (p1p6 results). If the isolated p15 viruses were incubated for an additional 24 h at 37 °C in cell-free growth medium, the percentage of gRNA dimerization reached 100% of WT (data not shown), indicating that the p15 mutation slowed down the kinetics of gRNA dimerization, rather than impose a fixed level of dimerization. Song *et al.* (253) had shown that the percentage of gRNA dimerization attains its maximal level a virus age of < 1h in WT. The present data suggest that the p15 mutation slowed down gRNA dimerization kinetics > 10 fold, if not about 50 fold.

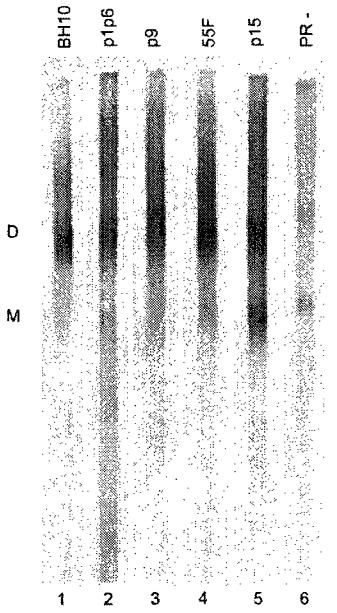


Fig. 3. 4. Dimerization level of viral RNA isolated from HIV-1_{HXB2} and virions mutated Pr55gag polyprotein sessile sites. Genomic RNAs extracted from the respective virions were electrophoresed on a 1% non-denaturing agarose gel and analyzed by Northern blotting. The representative lanes contain viral gRNA isolated from one 35 mm tissue culture dish, and the autoradiographic exposure times varied from 20 min to 6 h. Mutants are defined in Fig.1. D: dimer. M: monomer. BH10 (HXB2) gRNA samples were 77 % ± 0.5 % dimeric (n = 29) and PR^- gRNA samples were 46 % ± 2.5 % dimeric (n = 3). The gRNA dimerization level is independent of the amount of gRNA electrophoresed or of the concentration of DNA used in

transfections (25-fold range of gRNA/proviral DNA concentrations tested with BH10 and an HIV-1_{HXB2} mutant bearing an inactivated dimerization initiations site (253) (not shown).

Genomic RNA packaging seemed unchanged by the mutations, except that mutant 55F reduced the amount of gRNA packaged per unit of CAp24 (Table 1), as if premature apparition of NCp7 impaired this process. RT packaging, in all mutants was almost equal to wild type. Virus stability was reduced by the p1p6, p15 and 55F mutations but unaffected by the p9 mutation (Table 1).

Full genomic RNA dimerization despite 75% unprocessed Pr55gag polyproteins

The first step in viral protein processing starts with the release of the viral protease from Gag-Pol polyprotein. Since PR is a dimeric enzyme, dimerization of Gag-Pol is required for enzyme activation (144,331). Retroviral proteases belongs to the aspartyl protease family, and uses two opposed catalytic aspartic acid residues each embedded in a motif on separate molecules to catalyze a peptide bond in a target protein (332,333). Therefore, the active enzymes form a dimer with catalytic site being in the center of the molecule and substrate binding site on the periphery (334,335). In each scissile site, the enzyme interacts with seven successive amino acid residues of the polyprotein substrate in order to carry its catalytic action between P1 and P1' amino acids (the P1 position is the amino acid immediately upstream of the scissile bond, and the P1' position is the amino acid immediately downstream of the scissile bond). Even though the amino acids flanking the scissile site are generally hydrophobic, there is no consensus sequence, no doubt in part because few if any Pr55gag processing sites need to be processed at maximal rates (56,336). Instead, accessibility or exposure of the cleavage site in suitable vicinity appears to govern the efficacy and consequently the rate of cleavage in each scissile site. This broad range of cleavage efficiencies lead to stepwise and ordered processing of the Pr55gag

polyprotein (207). Blocking the enzyme's activity by using protease inhibitors is among the first drugs used to control HIV-1 infection (337). Subtle defects in processing, achieved by suboptimal anti protease drug, correlate with profound deficits in infectivity, altered virion morphology and a reduced capacity to carry out DNA synthesis (338). Different groups have studied the enzyme at the molecular level. The effect of partial inactivation of the protease (i.e. of the number of WT proteases needed) has been studied to some extent (317,338-340) but the number of proteases needed for maturation of gRNA of HIV-1 is not very well documented.

The mutations Thr26Ser and Ala28Ser, in the protease active site, reduce 4-fold and 50-fold, respectively, the catalytic activity (k_{cat}) of recombinant HIV-1 protease against a decapeptide that mimics the protease-reverse-transcriptase cleavage site (317). We inserted these mutations in the BH10 provirus to produce mutants 4X and 50X, respectively. The effect of these mutations in isolated viruses confirms the *in vitro* results (Fig. 2C, lanes 10 and 11). Namely, the proportion of fully processed capsid protein is 10 %, 90 % and > 97 % in 50X, 4X , and BH10 virions, respectively (Fig. 2C), consistent with results of Rosé *et al.* (317). In order to delineate the amount of viral protease activity needed to produce mature virus 48 hours post transfection, we also studied, by cotransfection, the effect of various ratios of BH10 and PR- plasmids (Table 2).

	Construct name[a]	BH10 in Cotransfection	PR- in Cotransfection	Specific protease activity[b]	Gag Processing[c]	gRNA dimerization[c]
1	HXB2	100	0	100%	100%	100
2	PR-	0	100	0	0	60 ± 3
3	4X	-	-	25%	90 ± 2	97 ± 2
4	50X	-	-	2%	11 ± 3	69 ± 2
5	PR 1%	10	90	1%	4 ± 3	59 ± 1
6	PR 5%	22	78	5%	24 ± 6	92 ± 2
7	PR 10%	32	68	10%	43 ± 4	94 ± 2
8	PR 15%	39	61	15%	48 ± 4	94 ± 3
9	PR 20%	45	55	20%	74 ± 8	95 ± 2
10	PR 25%	50	50	25%	79 ± 7	96 ± 2
11	PR 40%	63	37	40%	87 ± 6	100 ± 2

Table 3. 2. HIV-1_{HXB2} and protease active

site inactivated provirus (Pr-) are used in proportionally calculated mixture to produce assumed protease activity in derived virions.

[a] Mutants are explained in the results and discussion section. Columns two and three show the proportion of wild type and active site inactivated provirus (Pr-) in each cotransfection respectively.
[b] Numbers describe the assumed protease activity in each cotransfection as explained in results and discussion.
[c] For each mutant, the error bars designate the standard error of 2 to 4 independent experiments for gRNA dimerization (3 on average), 2 to 4 independent experiments, for gRNA packaging (3 on average), and 2 independent experiments for Pr55gag processing. The values for HXB2 are arbitrarily set at 100, and the values for the mutants are expressed as % of wild-type level. Genomic RNA dimerization numbers were obtained by densitometric analysis.

WT Pr160gag-pol and Pr160gag-pol mutated at Asp25 appear to have identical affinities for each other, and to bind randomly to each other when mutant and WT Pr160gag-pol are introduced in cells by cotransfection (332). The specific protease activity (i.e. per unit of protease molecules) yielded by a BH10/PR- cotransfection is therefore given by the expected proportion of BH10 protease dimers: for example, a 1 to 1 BH10/PR- cotransfection would yield a specific protease activity that is 25 % of WT. The various plasmid combinations, in the cotransfection experiments, were named according to the expected protease activity (PR 1%, PR 5 %, PR 10%, etc. in Fig. 2C, Table 2, Figs. 5a and 5b). We assumed that the co-transfected proviruses had equal chances to enter HeLa cells, that the proteolytic activity of uncleaved Pr160gag-pol on Pr55gag was for all practical purposes negligible compared to that of the mature protease because the initial steps in the processing of the Pr160gag-pol precursor produced *in vitro* are intramolecular, and not extramolecular, cleavages (312,341) and that the affinity of Pr160gag-pol for another one is unchanged by the presence of an aspartate or an arginine at position 25 of its viral protease. Numerous crystal structures of the mature, 99-amino-acid PR have been described, and several groups have conducted exhaustive genetic and biophysical analyses of the mature dimer (335,341). Western blotting experiments indicate that

in the PR 1%, PR 5% and PR 10% cotransfections, the Pr55gag of the isolated viruses was processed to a roughly 4%, 24 % and 43 % level (Fig. 2C, lanes 2-4). The 9% Pr55gag processing seen with the PR 1% cotransfection most likely indicates that either the Asp25/Arg25 protease heterodimer is not completely inactive, or 1% protease activity for 48 hours is enough to process almost 5% of viral Pr55gag and Pr160gag-pol polyproteins. The amount of Pr55gag processing in 4X mutant and 25% cotransfection which theoretically expected to be identical shows different numbers of 90% and 79% respectively (Fig2C, compare lanes 7 and 10). A difference was also seen in the gRNA dimerization yield of these two preparations. The observed difference may reflect a greater protease activity for 4X mutant *in vivo* compared to *in vitro* experiments or may reflect a possible protease activity for heterodimeric protease enzyme molecules composed of BH10 and PR- heterodimers in the *in vivo* situation. The percentage of gRNA dimerization and pr55gag protein processing follow the same trend when plotted against specific protease activity (Fig 5B) and both parameters show rapid rise with small increase of protease activity from practically 0 in Pr- mutant to 25% in 5% protease active co-transfection experiments. Dimerization of gRNA increases form 60% in PR- mutant to 92% and Pr55gag processing rises from zero to one quarter of total Pr55gag and Pr160gag-pol polyproteins in the presence of 5% protease activity (Figs 2C,5A and 5B). In any protease activity, the amount of gRNA dimerization is far ahead of Pr55gag processing and reaches sooner to its peak.

The percentage of gRNA dimerization was 97%, 92%, 69%, and 59% of WT in 4X, PR 5%, 50X and PR 1% virions, respectively (Fig. 5a and Table 2). Thus, gRNA dimerization as observed in 0 to 48 h viruses is not significantly affected by a 7-fold reduction in proteolytic processing, but is seriously affected by the slightly larger reduction in proteolytic processing seen in 50X viruses (Fig. 2c, Fig. 5a, Fig. 5b, and Table 2). Thus, about 600 processed Pr55gag

(out of 4000 in WT) may be sufficient for the purpose of achieving WT gRNA dimerization in 0-48 h viruses. The additional number of NC present in WT viruses may be required for some other of its varied functions, or to achieve full gRNA dimerization in a shorter time. The PR 5% results indicate that 25% Pr55gag processing in 0-48 h viruses produced a greater gRNA dimerization yield than 80 % Pr55gag processing in 0-5 min viruses (253). Thus gRNA dimerization yield is, naturally, a function of the number of active viral proteins available in the virus, and of the time they have had to act on gRNA.

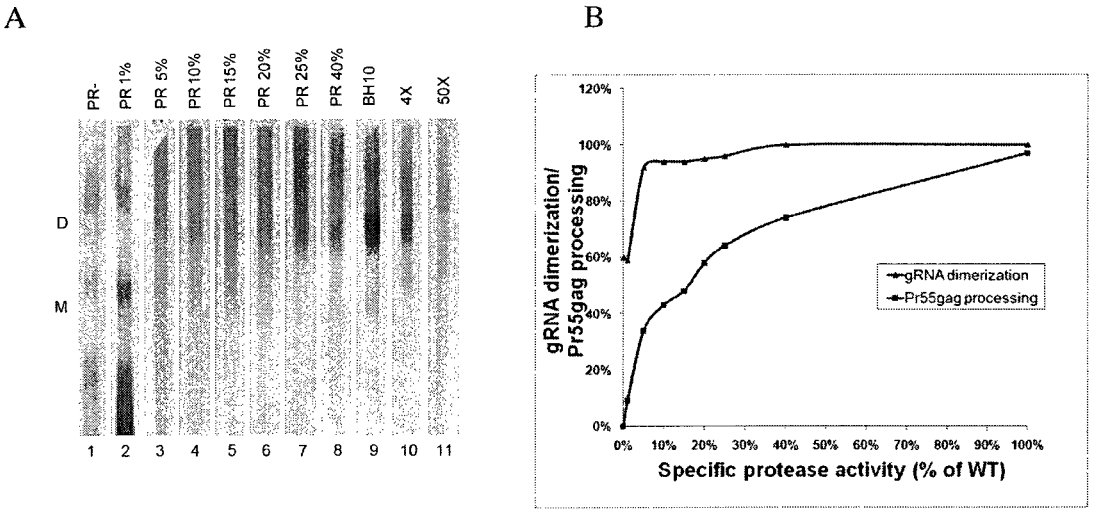


Fig. 3. 5. A) Dimerization level of viral RNA isolated from protease weakening mutants and HIV-1_{HXB2} with Pr- cotransfections. Experimental conditions as in Fig. 3. Exposure times varied from 15 min to 4 h. B) Dimerization level and Pr55gag processing of virions derived from HIV-1_{HXB2} and Protease inactivated plasmids plotted against assumed protease activity. X axes represents assumed protease activity and Y axes shows gRNA dimerization and Pr55gag processing.

In 50X viruses, Pr55gag processing as well as gRNA dimerization were intermediate between that seen in PR 1% virions and PR 5% virions. Though this is consistent with the 2 % protease activity seen *in vitro* on a model decapeptide (342), detailed comparisons with the cotransfection experiments should be avoided because nothing guarantees that the 50X mutant impairs equally each Pr55gag and Pr160gar-pol cleavage site.

The viruses PR 5%, PR 10%, PR 15%, PR20%, PR25% , PR40% and 4X show wild type like gRNA dimerization after being incubated at 37 °C in cell-free growth medium for 24 hours. Dimerization of 50X is not changed after incubation (data not shown).

2000 disabled NC (half the complement) do not impair gRNA dimerization in isolated viruses

The NCp7 protein is a sequence of 55 amino acid residues, 15 of which are highly basic and only four highly acidic. Like all lentiviral NC, it possesses two zinc containing motifs (often called zinc fingers) of the form **Cys**X_2**Cys**X_4**His**X_4**Cys,** where X = variable or conservatively substituted amino acid residue (30,267-269,294). In our previous study on NCp7 (30) we concluded that the amino-terminus, the proximal zinc finger, the linker, and the distal zinc finger of NC each contribute roughly equally to efficient gRNA dimerization, and each segment of NC plays a role, direct or not, in gRNA dimerization, gRNA packaging, Pr55gag processing, Gag-Pol incorporation, viral infectivity and virus stability. Another event that occurs during particle assembly is the incorporation of tRNA molecules into virions. These serve as the primer for the initiation of reverse transcription (225). It has been demonstrated from cell-free assays that HIV-1 NC, both in the precursor (Pr55Gag) and mature (NCp7) forms, enables the stable annealing of its cognate tRNA to gRNA sequences, which entails the unfolding of 15 nucleotides of the tRNA

(61,146,228,343-345). The conversion of gRNA to full-length dsDNA is an essential early step in retroviral replication and RT must perform its functions in the context of gRNA being completely coated with NC (346). It has also been observed that NC stabilizes nascent vDNA in the cytoplasm of infected cell (170-172,347). Observations from infection experiments performed with certain NC mutant viruses have provided indirect evidence for the involvement of NC in integration of full-length vDNA into the chromosomal DNA of the infected cell (127). Composite virus particles derived from equimolar cotransfection of 293 cells with both wild type and mutant proviruses harboring mutations in either or both zinc fingers of NCp7 encapsidate gRNA with a similar efficiency and specificity as wild type particles, and can replicate (302).

The immature virus is composed of about 5000 Pr55gag polyproteins (184) or at least 1500 (Zhu *et al.*, 2003). After Pr55gag processing, less than one third of the freed CA proteins (~1500 free CA) contributes to the mature viral core (184,194,348-350). Because all NC molecules (~5000) are likely to be confined in the core, there is a possibility that a full complement of NC molecules is needed to obtain full gRNA dimerization.

The importance of the number of WT NC in HIV-1 is one that needs clarification. In the first section of this paper, we have shown that NCp7 is not needed for gRNA dimerization (p9 results) and that the presence of 25% NCp7 and 75% NCp15 (p1p6 results) fully compensates for the slow dimerization seen in the absence of NCp7 and NCp9 (p15 result). But in each of these experiments there was a full complement of essentially wild-type NC in the virus. In the 2nd section of the paper, we showed that 15 % processed Pr55gag, i.e. a 15 % complement of NCp15/NCp9/NCp7, is enough to achieve full gRNA dimerization and is undisturbed by the 85% of NC still part of Pr55gag. The question we pose now is whether dimerization would remain undisturbed if the complement was disabled NC.

To answer this question, we performed cotransfection experiments in which mutant proviruses altered in different segments of NC protein were mixed with equal amount of WT. 48 hours post transfection, the supernatant was collected, gRNA was extracted from purified viruses, gRNA dimerization was studied by native northern blot (Fig 6). The NC mutants chosen were highly disabling via mutations in the proximal zinc finger (3EF1; ΔF1), the linker (S3E; FVI) or the second zinc finger (ΔF2) (30). But in each case, the severely debilitating effect of the NC mutations on gRNA dimerization was fully rescued when the viruses contained a 50% complement of wild-type NC. These mutations are described and studied before in our laboratory (30). Deletion of zinc finger 1 (ΔF1) and finger 2 (ΔF2) reduced almost equally the gRNA dimerization to approximately 72% and 76% of wild type respectively. Pr55gag processing in ΔF1 was 54% of wild type but ΔF2 had less effect on Pr55gag processing by reducing to 84% of wild type (30). The decreased amount of dimerization observed in ΔF1 compared to ΔF2 attributed to the effect of the inserted mutant rather than unprocessed Pr55gag proteins (30). Both ΔF1 and ΔF2 mutations when co transfected with equal amount of wild type provirus, showed wild type amount of gRNA dimerization, while the amount of unprocessed Pr55gag was not been changed (Fig.2D and Fig.3 in Kafaie et al., 2008). We observed the same trend in 3EF1 mutant where three basic residues in finger 1 (K14K20R26) replaced by three glutamic acid (E) residues(30). 3EF1 reduced processing of Pr55gag protein to 80% of wild type which was increased to 88% when cotransfected with wild type BH10 in equimolar proportion. We previously have introduced S3E and FVI mutants in linker of NC by mutating ArgLysLys34 into PheValIle (FVI) or GluGluGlu (S3E). These two mutants reduced the charge of the linker from +4 to +1 and -2 respectively. The effect of FVI on gRNA dimerization is comparable to the reductions noticed in ΔF1 and ΔF2, but S3E has the strongest effect for a single mutant reported

so far by reducing gRNA dimerization to 45% of wild type (30). Our cotransfection studies showed that gRNA dimerization in pelleted viruses from equimolar transfection of BH10 and any of these two proviruses is almost at the same level as the WT (Table 3 and Fig.6).

Western blot studies with anti CA (p24) antibody showed no significant change in Pr55gag polyprotein processing in NC mutant and NC mutant in cotransfection with BH10 in equimolar proportion (Fig.2D, Table 3, (30)). In other words, increasing wild type NC protein in viral vicinity to 50% is enough to increase gRNA dimerization in selected mutants regardless of their structural and/or electrical deficiency to the level of wild type. The ability of wild type Pr55gag protein to rescue infectious incompetent variant may also have therapeutic implication. In this set of experiments we assumed that the co-transfected proviruses had equal chances to enter HeLa cells, that mutated and wild type Pr55gag proteins had equally translated and incorporated in viral assembly resulting a uniform hybrid virus production.

	Transfection	Dimerization	Pr55gag Processing		Transfection	Dimerization	Pr55gag Processing
1	BH10	100	100	7	ΔF1	72.5 ± 3	54
2	PR-	61 ± 2.6	0	8	ΔF1+ (1:1)	97 ± 2.5	58
3	S3E	45 ± 2	64	9	ΔF2	76 ± 1.5	84
4	S3E+ (1:1)	100 ± 1.5	68	10	ΔF2+ (1:1)	95 ± 2.1	90
5	3EF1	68 ± 4	80	11	FVI	74 ± 2	97
6	3EF1+ (1:1)	100 ± 1.5	88	12	FVI+ (1:1)	98 ± 2	95

Table 3. 3. Dimerization and Pr55gag processing of mutants.
Mutants in different segments of NCp7 in separate (Mutant) and as 1:1 proportions with HIV-1_{HXB2} (Mutant+) are used to do co-transfection rescue experiments. Dimerization of gRNA can be rescued in all mutants harboring a mutation in NCp7 without any significant change in their respective Pr55gag processing.

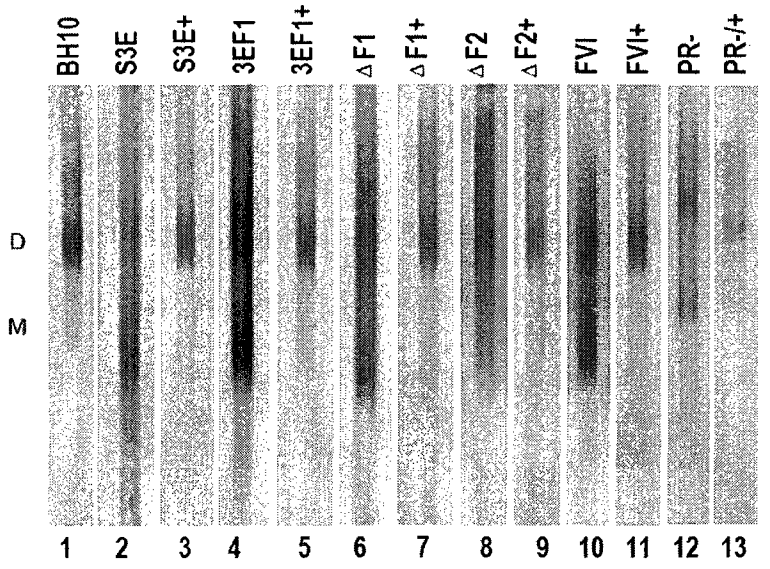


Fig. 3. 6. Dimerization level of viral RNA isolated from 1:1 co-transfections of HIV-1_{HXB2} and several NC mutated plasmids. Experimental conditions as in Fig. 5. Exposure times varied from 15 min to 16 h.

3.4. Conclusion

Genomic dimerization is an important step in the viral life cycle since impairing this pathway results in reduced genome packaging as well as defective viral structure. Newly released viruses lack gRNA dimers which then accumulates in < 1hr old wild type viruses. The genomic RNA dimerization process is associated with Pr55gag processing. In our previous work we studied thoroughly the effect of N-terminus, finger 1 and 2 and Linker regions of NCp7 protein in gRNA dimerization and other viral life cycle. In this paper we first focused on the effect of intermediate proteins derived from Pr55gag processing on gRNA dimerization and packaging, and viral stability and RT incorporation to the virus. We found that Pr55gag processing to NCp9 is sufficient enough to bring gRNA dimerization to wild type level even though mutant p9 lacks infectivity. In the second section of the paper, we changed protease activity, via mutational and cotransfection studies, to understand the minimal requirements for protease activity in gRNA dimerization. We concluded

that 5% enzyme activity is enough to yield 90% of wild type level gRNA dimerization despite 75% unprocessed Pr55gag molecules. The remaining unprocessed Pr55gag molecules did not interfere with gRNA dimerization. In the third section of the paper, we conducted cotransfection studies using equal proportions of BH10 and NC mutated proviral DNAs to study the ability of wild type NC protein to chaperone gRNA dimerization. We found that regardless of the site of mutation in the NC sequence (e.g. the change in electrical charge), if 50 % of the NC is wild-type, this is sufficient to bring gRNA dimerization to a level identical to wild type. All together these data shed tight on the importance of Pr55gag processing in the viral life cycle and can partially explain viral drug resistance observed in clinical encounters.

3.5. Materials and Methods

Plasmid construction. Proviral vector pSVC21.BH10 encodes a HIV-1_{HXB2} cDNA clone. Mutant proviral vectors, except p15, were constructed from pSVC21.BH10 by PCR mutagenesis, using primers described in Table 4. The nucleotides positions are based on the sequence of HIV-1 gRNA. To prepare mutants in Gag-p55 scissile sites (55F, p9 and p1/p6) and Protease weakening (4X and 50X) mutants, a DNA fragment extending from Apa I to Bcl I restriction sites was synthesized with desired mutations by PCR, and ligated into pSVC21.BH10. p15 was constructed using mutant p9 as template in PCR and the same primers used to construct p1p6 mutant. Then, the constructed DNA fragment from Apa I to Bcl I restriction sites (Amersham) was ligated into pSVC21.BH10. To prepare CA mutant, restriction sites Nar I and Spe I were used with related primers depicted in table 4, and the DNA was introduced into pSVC21.BH10 by ligase (Amersham). After mutagenesis and ligation, all mutated DNA fragments produced by PCR were sequenced (ACGT Inc., Toronto) and confirmed to contain the correct mutations.

Construct name	Primer (all primers are sense)[a]
P9	5'gattgtactgagagacaggcttctttttagggaagatctggccttcc
P1p6	5'cctacaagggaaggccagggaatcttcttcagagcagaccagagccaac
55F	5'gattgtactgagagacaggctttttttttagggaagatctggccttcc
4X	5'ctaaaggaagctctattagattcaggagcagatgatacag
50X	5'ggaagctctattagatacaggatcagatgatacagtattagaag

Table 3. 4. Primers used to introduce intended mutations in HIV-1_{HXB2}

[a]All primers were synthesized by ACGT Inc. (Toronto).

Cell culture and transfections. HeLa Cells were cultured at 37 °C in a medium consisting of Dulbecco's modified Eagle's medium (DMEM), 10% fetal calf serum, ampicillin and streptomycin (Invitrogen). The PolyFect transfection reagent (Qiagen) was used to transfect 9 μg of proviral DNA into 50% to 70% confluent HeLa cells in 100- by 20-mm petri dishes containing 10 ml of culture medium.

Infectivity assay. Mutant proviruses and the parental BH10 provirus were independently transfected into HeLa cells. Virus-containing supernatants were collected 48 h post-transfection and passed through 0.2 μm pore-size cellulose acetate filters to remove the cells. The CAp24 content of these clarified supernatants was measured using an ELISA kit (Vironostika HIV-1 Antigen, Biomerieux). Equal amount of the supernatants (10 ng of CAp24 content) were used to infect equal numbers of MT2 cells (6×10^6 cells in 10 ml of RPMI 1640 medium, 10% fetal calf serum, ampicillin and streptomycin (Invitrogen), per petri dish). In the human T-cell line MT2, only a short time lag separates infection from viral replication. After 2 h, cells were washed twice to remove unbound viruses and were then maintained in serum-supplemented medium. On every other day, cells were diluted 1 in 2 into fresh medium and the RT activity in the supernatant of the removed medium was determined. RT activity measurements were made over a period of 14 days.

RT activity. The exogenous (oligo (dT) directed) RT activity was measured by adding 40 µl of RT cocktail (60 mM Tris–HCl [pH 7.9], 180 mM KCl, 6 mM $MgCl_2$, 6 mM dithiothreitol, 0.6 mM EGTA, 0.12% Triton X-100, 6 µg/ml oligo (dT), 12 µg/ml poly(rA), 0.05 mM ^{3}H dTTP) to a 10 µl sample. After incubation for 2 h at 37 °C, the reaction was stopped with cold 10% TCA (150 µl per well), and precipitated for 30 min at 4 °C. The precipitate was blotted, washed and scintillation counted.

Virus purification and isolation of HIV-1 viral RNA. Filtered virus-containing supernatants were centrifuged (SW41 rotor, 35 000 rpm, 4 °C, 1 h), through a 2 ml 20% (w/v) sucrose cushion in phosphate-buffered saline (PBS). The virus pellet was dissolved in 400 µl sterile lysis buffer (50 mM Tris (pH7.4), 50 mM NaCl, 10 mM EDTA, 1% (w/v) SDS, 50 µg tRNA per ml, and 100 µg proteinase K per ml), and extracted twice at 4 °C with an equal volume of buffer-saturated phenol-chloroform-isoamylalcohol (25:24:1) (Invitrogen). The aqueous phase was precipitated overnight at – 80 °C with 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of 95% ethanol, and centrifuged at 14,000 rpm in an Eppendorf 5145 micro centrifuge at 4 °C for 30 min. The gRNA pellet was rinsed with 70% ethanol, and dissolved in 10 µl buffer S (10 mM Tris (pH 7.5), 100 mM NaCl, 10 mM EDTA and 1% SDS) (30).

Electrophoretic analysis of HIV-1 gRNA. The gRNA was electrophoresed under non-denaturing conditions and identified by Northern (RNA) blot analysis (30). Electrophoretic conditions were 4 V/cm for 4 h on a 1% (w/v) agarose gel in TBE2 (89 mM Tris, 89 mM Borate and 2 mM EDTA, pH 8.3) at 4 °C. After electrophoresis, the gel was heated at 65 °C for 30 min in 10% (w/v) formaldehyde, and the embedded RNAs were diffusion transferred to a Hybond N+ nylon membrane (Amersham). After drying at room temperature for 2 h, crosslinking (3000 j in a UV Stratalinker),

and prehybridization at 42 °C for 3 h in 6× SSPE (1× SSPE is 0.15 M NaCl, 10 mM NaH2PO4, and 1 mM EDTA pH 7.4, 50% (w/v) deionized formamide, 10% dextran sulfate, 1.5% SDS, 5× Denhardt's reagent, 100 μg/ml salmon sperm DNA, the membrane was hybridized overnight in prehybridization buffer devoid of Denhardt's reagent in a rotating hybridization oven at 42 °C to approximately 25 μCi of ^{35}S-labeled antisense RNA 636-296 (a 356-nt RNA that is the antisense of the 296 to 636 region of the HIV-1 genome prepared with the SP6 Megascript kit [Ambion]) (Laughrea and Jetté, 1996). This was followed by two 30 min washes in 1× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate]– 0.1% SDS at room temperature and 37 °C, and one 30 min wash in 0.2× SSC–0.1% SDS at 45 °C (246), exposure to a Kodak BioMax MR X-ray film, and densitometric analysis.

Densitometric analysis. The autoradiograms were scanned and analyzed with the NIH 1.6.3 program. Care was taken to scan variously exposed films to guard against over-exposed or under-exposed bands or spots. The monomer and dimer bands were considered of equal width; that width was approximately twice the vertical size of the D and M letters used to indicate dimers and monomers in the relevant figures. Material located elsewhere in the gels was not taken into account in the calculation of the percentage of dimers. The diffuse character of many bands may reflect conformational diversity among the gRNA molecules. It is not due to poor resolution of the gels because heat denatured gRNAs formed a sharp band at the monomer position (not shown). The RNA bands seen here are not more diffuse, and sometimes sharper, than what was seen in earlier studies by us and others (220,223,246,258,260,289).

Genomic RNA packaging. The amount of gRNA per unit CAp24 of virus was quantified by hybridization with antisense RNA 636–296 using a dot–blot assay. Virus pellets were resuspended in

400 µl of Trizol LS reagent (Invitrogen), and incubated at 30 °C for 5 min. 100 µl of chloroform was added, followed by shaking for 15 seconds and incubation at room temperature for 15 min. After centrifugation (12,000 ×*g*, 15 min, 4 °C), the colorless aqueous upper phase was mixed with 250 µl of isopropyl alcohol, incubated at room temperature for 10 min and centrifuged again. The precipitated RNA was washed once with 500 µl of 70% ethanol, pelleted (7500 ×*g*, 5 min, 4 °C), air-dried, dissolved in 10 µl RNase-free water and stored at – 20 °C. Serial 10-fold dilutions of wild type RNA samples, normalized for input virion CAp24, were used to construct a standard curve. 29 µl of buffer F (100% deionized formamide, 20 µl; 20× SSC, 2 µl; 37% formaldehyde, 7 µl) was added to each sample, followed by incubation (68 °C, 15 min) and chilling on ice. After adding 78 µl of 20× SSC buffer, samples were vacuum-suction transferred to a Hybond N^+ nylon membrane (Amersham) sandwiched within a Hybri-Dot filtration manifold (Bethesda Research Laboratories). The wells were washed twice with 1 ml of 10× SSC, and suction continued for a further 5 min to dry the membrane. The membrane was removed, dried for 4 h, cross-linked, pre-hybridized, hybridized, autoradiographed and scanned as for Northern blot analysis (above). To confirm the scans, each individual spot of the nylon membrane was excised and scintillation counted.

Virus stability and RT packaging. 1.2 ml of filtered virus-containing supernatant was pelleted through a 0.3 ml 20% sucrose cushion in the TL-100 Beckman ultracentrifuge (TLA 55 rotor, 45,000 rpm, 1 h, 4 °C). The virus pellet was dissolved in 10 µl of PBS and its CAp24 content was measured using an ELISA kit (Vironostika HIV-1 Antigen, Biomerieux). The CA content of the purified viruses divided by the CA content of the 48 h culture supernatant, relative to the ratio found in WT, was taken as a measure of virus stability (Wang *et al.*, 2002; 2004). The ratio pellet/supernatant was 0.43 ± 0.07 in WT (this was taken to mean 100 in Table 1). The exogenous

RT activity of the pelleted viruses divided by their CA content, relative to the ratio found in WT, was interpreted as RT packaging.

Pr55gag proteolytic maturation. At 48 h post-transfection, cells were lysed in ice-cold NP-40 containing buffer (100 mM NaCl, 10 mM Tris, 1 mM EDTA, 0.5% NP-40, and protease inhibitor cocktail [Roche]). Supernatants were cleared by centrifugation at 3000 ×*g* and filtered (0.22 μm). Viruses were then concentrated through a sucrose cushion by ultracentrifugation. Equal amounts of viruses (judged by CAp24-ELISA) were lysed in the above buffer and subjected to SDS-polyacrylamide gel electrophoresis. Viral proteins were detected by Western blotting using a rabbit anti-capsid antibody (ABT-Trinity Biotechnology, CA, USA) that recognizes Pr55gag, CAp24, and CA-containing partially cleaved Gag proteins, a polyclonal rabbit anti-NC antibody (gift from Gorelick) and an enhanced chemiluminescence Western blot detection kit (Amersham). The signals for CA-containing proteins were quantified by densitometric scanning and analyzed with the NIH 1.6.3 program. The signal obtained from the CAp24 band was divided by the total signal obtained from all CA-containing proteins, to calculate Pr55gag processing level.

Note. *P* values were measured using the Student's *t* test.

Chapter 4
Conclusion and Contributions of Original Knowledge to the Field

4.1. Conclusion

The work presented below is a summary of the main findings contained in Chapters 2 and 3 of this thesis. The manuscript presented in chapter 2 has been published in refereed scientific journal (30) and chapter 3 is under preparation for submission.

Prior to the initiation of the projects presented in this thesis, there was little evidence about the role of NC protein in HIV-1 gRNA dimerization. Our lab has already shown that no gRNA dimers are detectable in newly released protease inactive (PR^-) HIV-1, but thermolabile and partially dimeric viral RNA dimers slowly accumulate with time (253). One or more maturation products of the viral polyproteins is required for fast or complete formation of thermolabile viral RNA dimers, and for refolding of thermolabile dimers into mature, thermostable dimers (253). Among these maturation products, a central role is attributed to the HIV-1 nucleocapsid protein (NC) for at least four reasons: 1) avian and murine NC proteins are implicated in the dimerization of avian and murine gRNA (118,250,254-256). 2) NC stimulates the *in vitro* dimerization of partial HIV-1 RNA transcripts containing the gRNA dimerization initiation site (131,146,251). 3) The HIV-1 NC stabilizes HIV-1 RNA dimers (251) and murine retroviral RNA dimers (91). 4) A mutation in NC can decrease HIV-1 gRNA dimerization as much as disruption of the dimerization initiation site (252).

The DIS is located in the apical loop of SL1, a 35 nucleotide-long stem-loop belonging to the 5' untranslated region of gRNA. This has stimulated studies on the role of NC in the dimerization of SL1 or molecules similar to SL1, with emphasis on: 1) the rearrangement of these molecules from a metastable kissing dimer to a stable extended duplex (261-266) and 2) the effect of NC mutations on the dimerization of a 39mer RNA similar to wild type (WT) SL1 but lengthened by two additional base-pairs for technical reasons (244,245). A role for the zinc fingers and the basic regions of NC was identified in *in vivo*, but there were two major ambiguities: 1) substituting completely scrambled N-

terminal or linker sequences for the WT sequence did not inhibit dimerization, as if nonspecific effects were studied under the in vitro conditions (245); 2) the two zinc fingers were jointly mutated, preventing an identification of the implicated finger(s) (244). No thorough study to this date had investigated the role of different segments of NC on HIV-1 genomic RNA and viral life cycle.

In vitro studies of the dimerization of partial HIV-1 RNA transcripts can be prone to artefacts (270). It is thus crucial to also investigate the effect of mutations on the properties of WT gRNA produced by viruses. Regarding the effect of NC mutations, gRNA dimerization was unaffected by a 75 % reduction of the number of basic residues in the N-terminus or the linker; and inactivating the C-terminal zinc finger inhibited gRNA dimerization as much as inactivating the DIS, but clearly less than inactivating the viral protease (252). Since mature NC is believed to be solely responsible for the increased dimerization seen in WT virions relative to PR^- virions, this differential effect suggests that NC segments other than the C-terminal zinc finger can modulate gRNA dimerization. Thus many issues were left unresolved, such as the role of the N-terminal zinc finger and its components (e.g. its hydrophobic patch (or cleft), its highly basic residues, its highly conserved histidine and glycines), the role of the corresponding components in the distal zinc finger, a deeper inquiry into whether the N-terminal and linker segments can modulate gRNA dimerization. Moreover, the NC mutations were studied in the NL4-3 isolate (252). HIV-1_{NL4-3} is less impaired by the neutralisation of basic NC residues than the HXB2-BH10 isolate (271,272), even though it is also a subtype B strain (HIV-1_{NL4-3} has a NY5 5' half and a HXB2-Lai 3' half) (273).

In this presented thesis, the impact of 40 NC mutations on gRNA dimerization was investigated. Thirty-eight were investigated for the first time; the remaining two were previously studied in HIV-1_{NL4-3} (252). The 40 mutations were engineered in the first four segments of NC from HIV-1_{HXB2}. They collectively involve 34 of their 49 residues (the six-residue C-terminal segment

was not mutated). In each investigated segment, relevant and dispensable residues were identified. The results reveal that the amino-terminus, the proximal zinc finger, the linker, and the distal zinc finger of NC each contributed roughly equally to efficient gRNA dimerization, the N-terminal and linker segments playing predominantly electrostatic and steric roles, respectively. Mutating the hydrophobic patch or the conserved glycines of either zinc finger was as disabling as deleting the corresponding finger; some mutations involving as little as three NC residues impaired gRNA dimerization more than inactivating the viral protease; the charge of NC could be increased or decreased by up to 18% without affecting gRNA dimerization.

NC mutations can also impair other late infection events such as gRNA packaging (51,124,130,239,274-277), virus assembly or stability (109,192,201,222,276,278-280,351), as well as packaging of Pr160gag-pol or reverse transcriptase activity (124,221). In this thesis, we investigated the effect of our NC mutations on genomic RNA packaging, virus stability, RT packaging, Pr55gag processing, as well as on viral replication.

Highlights of new results presented in this dissertation

Our main results can be briefly summarized as follows. 1) The amino-terminus, the proximal zinc finger, the linker, and the distal zinc finger of NC each contribute roughly equally to efficient gRNA dimerization. 2) The N-terminal and linker segments appear to play predominantly electrostatic and steric roles, respectively. 3) Mutating the hydrophobic patch or the conserved glycines of the proximal or the distal zinc finger is as disabling as deleting the corresponding finger, while mutating the CCHC motif of either finger, interchanging the zinc fingers, or replacing one zinc finger by a copy of the other one, has generally intermediate effects. 4) An acidic linker, or a joint mutation in the proximal finger and the linker, can impair gRNA dimerization more than an inactive viral protease. 5) The charge of NC can be increased or decreased by up to 18% without affecting

gRNA dimerization. 6) Each segment of NC plays a role, direct or not, in gRNA dimerization, gRNA packaging, and virus stability. 7) Sixteen mutations inhibited gRNA packaging more than 3-fold; eleven mutations inhibited gRNA packaging less than 3-fold, yet blocked viral replication. 8) Some mutations, involving no more than two or three NC residues, reduced virus stability 5 to 6-fold; this is not far from the 10-fold reduction previously seen as a result of jointly mutating 10 highly basic NC residues (279,280).

Regarding gRNA dimerization, the effects of 38 and 40 of the NC mutations were studied for the first time in HIV-1 and in HIV-1_{HXB2}, respectively. Regarding virus stability and RT packaging, the effects of all mutations were novel. Regarding gRNA packaging, the effects of 20 and 32 of the mutations were studied for the first time in HIV-1 and in HIV-1_{HXB2}, respectively. Each NC segment appeared similarly important for gRNA packaging.

NC is implicated in late infection events (e.g. gRNA dimerization, gRNA packaging, Pr55gag processing and virus stability, all studied in my thesis) and in early infection events such as reverse transcription of gRNA and proviral DNA integration into the host genome (138,172). Each of the 40 mutations impaired one or several of the late functions studied. However, mutations P31A and N+ had mild effects on late infection events and little impact on viral replication: they are thus unlikely to inhibit early infection events in HIV-1_{HXB2}. With the exception of borderline 2KAF2, the 38 other mutations probably inhibit both early and late infection events because their impact on viral replication was inordinately large relative to their effect on late infection events. For example, mutation NC2-1 blocked viral replication and had relatively mild effects on late infection events (Table 2), suggesting that it impairs early infection events, which it does (287,288).

During my works, I introduced S3E and FL mutants that are the only retroviral NC mutations known to inhibit gRNA dimerization more strongly than protease inactivation. This suggests that the

gRNA dimerization seen in PR^- virions is stimulated by unprocessed viral proteins. But none of the mutants reduced gRNA dimerization close to zero.

NC electrical charge. I also studied the effect of electrical charge of NC on HIV-1 virus. Genomic RNA dimerization was unaffected by altering the charge of NC by up to 18 % (removing or adding one or two highly basic residues); removing three positive charges was generally disabling (R7 and 3AF1, but not S3); removing ≥ four was always disabling. Subtracting two to three, or four to six, positive charges from NC reduced gRNA packaging an average of 2-fold (see R7, 3AF1, S3, FVI and 2KAF2) and 6.5-fold (see R7E, 3EF1, S3E, 2KEF2, and NL), respectively. Subtracting one positive charge was harmless in the single example studied (K38N). Mutation N+ improved virus stability.

New findings in Linker Region: Regarding gRNA dimerization, the role of the linker appears more steric than electrostatic: a neutral and polar linker (S3) was innocuous, but a shortened (ΔAP, Δlinker), lengthened (LL), hydrophobic (FVI) or negatively charged (S3E) linker was at least as disabling as deleting zinc fingers 1 or 2. Regarding gRNA packaging, a long linker (LL) was two to four times less disabling than a short linker (ΔAP, Δlinker). This suggests that specific finger-finger interactions are less crucial for gRNA packaging than for gRNA dimerization. I also showed that shortened (ΔAP, Δlinker) and negatively charged (S3E) linker decease viral stability considerably.

New findings in N-terminus. The role of the N-terminus appears to be more electrostatic than that of the linker, since reducing the charge of the N-terminus from +5 to +2 (R7) impaired gRNA dimerization, genomic RNA packaging and viral stability. More basic N-terminus region had no effect on genomic RNA dimerization and RT packaging, increased viral stability two times and deceased genomic RNA packaging four times.

Importance of Glycine pairs in finger 1 and 2. In human cellular nucleic acid binding proteins and most lentiviral NCs, $GlyX_2Gly$ immediately precedes the histidine of each zinc finger (269,293). Consistent with this high degree of conservation, we found that substituting alanines for the glycines, in any one of the zinc fingers, inhibited gRNA dimerization and packaging as strongly as deleting the finger.

Powerful aromatic residues. In each zinc finger of lentiviral NCs and human cellular nucleic acid binding proteins, an aromatic residue adjoins the first cysteine (269,293). The point mutations F16A, W37A and W37F inhibited gRNA dimerization and packaging as strongly as deleting the fingers to which they belonged.

Importance of zinc fingers. Deleting the N-terminal or the C-terminal zinc finger reduced gRNA dimerization similarly. Zinc finger interchange or replacement of one finger by a copy of the other inhibited gRNA dimerization approximately 60% as well as zinc finger deletion but had no or little impact on gRNA packaging in HIV-1_{HXB2}. We have shown that a CCHH motif in any of the two zinc fingers, or a CCCC motif in the distal finger, inhibited gRNA dimerization and packaging $\geq$ 70 % as well as deleting the corresponding finger. While mutation H23C inhibited gRNA packaging as strongly as deleting the N-terminal zinc finger, it had modest inhibitory effects on gRNA dimerization (~ 30 % of the effect of deleting the finger). Thus lentiviral gRNA dimerization requires a zinc coordination motif that clearly cannot be replaced by CCHH or CCCC at the distal position and by CCHH at the proximal position. Point mutations H23A and H44A reduced gRNA dimerization as much as F16A, W37A, or deleting the fingers to which they belonged. Mutations C28S and C36S inhibited gRNA dimerization less efficiently than H23A and H44A, but reduced gRNA packaging as strongly as these mutations.

Context-dependent phenotype of some mutations. We have compared the effect of 10 mutations on gRNA dimerization in HXB2 and NL4-3 virions produced by HeLa cells. We have found that mutations FVI, C49H and HC were more disabling in HIV-1_{HXB2} than in HIV-1_{NL4-3}, and that H23C, C28H, H44C, CH, NC1-1, NC2-2 and NC2-1 had comparable effects in the two isolates. Overall, the results indicate that HIV-1_{HXB2} may be more sensitive than HIV-1_{NL4-3} regarding the effect on gRNA dimerization of 25% of the mutations investigated. It is also noteworthy that HIV-1_{HXB2} gRNAs were 8 ± 2% less dimeric than HIV-1_{NL4-3} gRNAs.

Pr55gag processing and gRNA dimerization. Newly released viruses lack gRNA dimers which then accumulates in < 1hr old wild type viruses. The genomic RNA dimerization process is associated with Pr55gag processing. In chapter two of this thesis we studied thoroughly the effect of N-terminus, Finger 1 and 2 and Linker regions of NCp7 protein in gRNA dimerization and other viral life cycle. In chapter three, we first focused on the effect of intermediate proteins derived from Pr55gag processing on gRNA dimerization and packaging, and viral stability and RT incorporation to the virus. We found that Pr55gag processing to NCp9 is sufficient enough to bring gRNA dimerization to wild type level even though mutant p9 lacks infectivity. We also changed the protease activity, via mutational and cotransfection studies, to understand the minimal requirements for protease activity in gRNA dimerization. We concluded that 5% enzyme activity is enough to yield 90% of wild type level gRNA dimerization despite 75% unprocessed Pr55gag molecules. The remaining unprocessed Pr55gag molecules did not interfere with gRNA dimerization. In the last section of my works presented in this dissertation, we conducted co transfection studies using equal proportions of BH10 and NC mutated proviral DNAs to study the ability of wild type NC protein to chaperone gRNA dimerization. We found that regardless of the site of mutation in the NC sequence (e.g. the change in electrical charge), if

50 % of the NC is wild-type, this is sufficient to bring gRNA dimerization to a level identical to wild type. All together these data shed light on the importance of Pr55gag processing in the viral life cycle and can partially explain viral drug resistance observed in clinical encounters.

References

1. (WHO), W. H. O. (2006) AIDS Epidemic Update.
2. Emini, E. A. (ed) (2002) *The Human Immunodeficiency Virus*, Princeton University Press, Princeton
3. Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997) *Retroviral Virions and Genomes in Retroviruses*

Cold Spring Harbor Laboratory Press: Cold Spring Harbor,, NY

4. Kanabus, A. a. A., S., The Origins of HIV & the first cases of AIDS. AVERT. http://www.avert.org/origins.htm.
5. Letvin, N. L. (2006) *Nat Rev Immunol* **6**, 930-939
6. (1981) *MMWR Morb Mortal Wkly Rep* **30**, 305-308
7. (1982) *MMWR Morb Mortal Wkly Rep* **31**, 652-654
8. (CDC), C. f. D. C. (2008)
9. Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D., and Gallo, R. C. (1980) *Proc Natl Acad Sci U S A* **77**, 7415-7419
10. Uchiyama, T., Yodoi, J., Sagawa, K., Takatsuki, K., and Uchino, H. (1977) *Blood* **50**, 481-492
11. Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., and et al. (1986) *Nature* **321**, 10
12. AVERT. (2008) www.avert.org.
13. Eaton, L. A., and Kalichman, S. C. (2009) *J Assoc Nurses AIDS Care* **20**, 39-49
14. Gahir, S., Anger, G. J., Ibrahim, M., Read, S., and Piquette-Miller, M. (2009) *Can J Clin Pharmacol* **16**, e68-77
15. Mills, E., Cooper, C., Anema, A., and Guyatt, G. (2008) *HIV Med* **9**, 332-335
16. Halperin, D. T., and Bailey, R. C. (1999) *Lancet* **354**, 1813-1815
17. Janeway, C. A., Jr. (2001) *Microbes Infect* **3**, 1167-1171
18. http://www.niaid.nih.gov/factsheets/howhiv.htm, N. I. o. A. a. I. D. U. (2004)
19. Basavapathruni, A., and Anderson, K. S. (2007) *FASEB J* **21**, 3795-3808
20. Sluis-Cremer, N., and Tachedjian, G. (2008) *Virus Res* **134**, 147-156
21. Evering, T. H., and Markowitz, M. (2008) *Expert Opin Investig Drugs* **17**, 413-422
22. Emmelkamp, J. M., and Rockstroh, J. K. (2007) *Eur J Med Res* **12**, 409-417
23. Greenberg, M., Cammack, N., Salgo, M., and Smiley, L. (2004) *Rev Med Virol* **14**, 321-337
24. Little, K., and Surjadi, M. (2000) *J Assoc Nurses AIDS Care* **11**, 19-28
25. Izumi, T., Shirakawa, K., and Takaori-Kondo, A. (2008) *Mini Rev Med Chem* **8**, 231-238
26. Newman, R. M., and Johnson, W. E. (2007) *AIDS Rev* **9**, 114-125
27. Scarlata, S., and Carter, C. (2003) *Biochim Biophys Acta* **1614**, 62-72
28. Cullen, B. R. (1991) *FASEB J* **5**, 2361-2368
29. Cullen, B. R. (1991) *Adv Virus Res* **40**, 1-17
30. Kafaie, J., Song, R., Abrahamyan, L., Mouland, A. J., and Laughrea, M. (2008) *Virology* **375**, 592-610
31. Craven, R. C., and Parent, L. J. (1996) *Curr Top Microbiol Immunol* **214**, 65-94

32. Spearman, P., Wang, J. J., Vander Heyden, N., and Ratner, L. (1994) *J Virol* **68**, 3232-3242
33. Zhou, W., Parent, L. J., Wills, J. W., and Resh, M. D. (1994) *J Virol* **68**, 2556-2569
34. Freed, E. O., Orenstein, J. M., Buckler-White, A. J., and Martin, M. A. (1994) *J Virol* **68**, 5311-5320
35. Wang, C. T., Lai, H. Y., and Li, J. J. (1998) *J Virol* **72**, 7950-7959
36. Bukrinsky, M. I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., and Stevenson, M. (1993) *Nature* **365**, 666-669
37. Gallay, P., Swingler, S., Song, J., Bushman, F., and Trono, D. (1995) *Cell* **83**, 569-576
38. von Schwedler, U., Kornbluth, R. S., and Trono, D. (1994) *Proc Natl Acad Sci U S A* **91**, 6992-6996
39. Wang, C. T., Yang, A. H., and Chiang, C. C. (2000) *J Med Virol* **61**, 423-432
40. Kiernan, R. E., Ono, A., Englund, G., and Freed, E. O. (1998) *J Virol* **72**, 4116-4126
41. Reicin, A. S., Ohagen, A., Yin, L., Hoglund, S., and Goff, S. P. (1996) *J Virol* **70**, 8645-8652
42. Haseltine, W. A. (1991) *FASEB J* **5**, 2349-2360
43. Dorfman, T., Bukovsky, A., Ohagen, A., Hoglund, S., and Gottlinger, H. G. (1994) *J Virol* **68**, 8180-8187
44. Gamble, T. R., Yoo, S., Vajdos, F. F., von Schwedler, U. K., Worthylake, D. K., Wang, H., McCutcheon, J. P., Sundquist, W. I., and Hill, C. P. (1997) *Science* **278**, 849-853
45. Zhang, W. H., Hockley, D. J., Nermut, M. V., Morikawa, Y., and Jones, I. M. (1996) *J Gen Virol* **77 (Pt 4)**, 743-751
46. Wang, C. T., and Barklis, E. (1993) *J Virol* **67**, 4264-4273
47. Darlix, J. L., Cristofari, G., Rau, M., Pechoux, C., Berthoux, L., and Roques, B. (2000) *Adv Pharmacol* **48**, 345-372
48. Huang, M., Orenstein, J. M., Martin, M. A., and Freed, E. O. (1995) *J Virol* **69**, 6810-6818
49. Yu, X. F., Dawson, L., Tian, C. J., Flexner, C., and Dettenhofer, M. (1998) *J Virol* **72**, 3412-3417
50. Pettit, S. C., Moody, M. D., Wehbie, R. S., Kaplan, A. H., Nantermet, P. V., Klein, C. A., and Swanstrom, R. (1994) *J Virol* **68**, 8017-8027
51. Zhang, Y., and Barklis, E. (1995) *J Virol* **69**, 5716-5722
52. Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M., and Kingsman, A. J. (1988) *Cell* **55**, 1159-1169
53. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988) *Nature* **331**, 280-283
54. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., and Sigal, I. S. (1988) *Proc Natl Acad Sci U S A* **85**, 4686-4690
55. Park, J., and Morrow, C. D. (1993) *Virology* **194**, 843-850
56. Oroszlan, S., and Luftig, R. B. (1990) *Curr Top Microbiol Immunol* **157**, 153-185
57. Miller, M., Jaskolski, M., Rao, J. K., Leis, J., and Wlodawer, A. (1989) *Nature* **337**, 576-579
58. Katoh, I., Ikawa, Y., and Yoshinaka, Y. (1989) *J Virol* **63**, 2226-2232
59. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. (1989) *Science* **245**, 616-621
60. Goff, S. P. (1990) *J Acquir Immune Defic Syndr* **3**, 817-831

61. Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998) *Science* **282**, 1669-1675
62. Bukrinsky, M. I., Sharova, N., Dempsey, M. P., Stanwick, T. L., Bukrinskaya, A. G., Haggerty, S., and Stevenson, M. (1992) *Proc Natl Acad Sci U S A* **89**, 6580-6584
63. Katz, R. A., and Skalka, A. M. (1994) *Annu Rev Biochem* **63**, 133-173
64. Engelman, A., Englund, G., Orenstein, J. M., Martin, M. A., and Craigie, R. (1995) *J Virol* **69**, 2729-2736
65. Shin, C. G., Taddeo, B., Haseltine, W. A., and Farnet, C. M. (1994) *J Virol* **68**, 1633-1642
66. Berger, E. A., and Alkhatib, G. (2007) *Eur J Med Res* **12**, 403-407
67. Markovic, I., and Clouse, K. A. (2004) *Curr HIV Res* **2**, 223-234
68. Altmeyer, R. (2004) *Curr Pharm Des* **10**, 3701-3712
69. Arrigo, S. J., and Chen, I. S. (1991) *Genes Dev* **5**, 808-819
70. Chang, D. D., and Sharp, P. A. (1989) *Cell* **59**, 789-795
71. Cullen, B. R. (1992) *Microbiol Rev* **56**, 375-394
72. Kingsman, S. M., and Kingsman, A. J. (1996) *Eur J Biochem* **240**, 491-507
73. Laspia, M. F., Rice, A. P., and Mathews, M. B. (1989) *Cell* **59**, 283-292
74. Garcia, J. V., and Miller, A. D. (1991) *Nature* **350**, 508-511
75. Schwartz, O., Marechal, V., Le Gall, S., Lemonnier, F., and Heard, J. M. (1996) *Nat Med* **2**, 338-342
76. Trono, D. (1995) *Cell* **82**, 189-192
77. Stevenson, M. (2003) *Nat Med* **9**, 853-860
78. Aguiar, R. S., and Peterlin, B. M. (2008) *Virus Res* **134**, 74-85
79. Wolf, D., and Goff, S. P. (2008) *Annu Rev Genet* **42**, 143-163
80. Gabuzda, D. H., Lawrence, K., Langhoff, E., Terwilliger, E., Dorfman, T., Haseltine, W. A., and Sodroski, J. (1992) *J Virol* **66**, 6489-6495
81. Miller, R. H., and Sarver, N. (1997) *Nat Med* **3**, 389-394
82. Willey, R. L., Maldarelli, F., Martin, M. A., and Strebel, K. (1992) *J Virol* **66**, 7193-7200
83. Schubert, U., Ferrer-Montiel, A. V., Oblatt-Montal, M., Henklein, P., Strebel, K., and Montal, M. (1996) *FEBS Lett* **398**, 12-18
84. Paxton, W., Connor, R. I., and Landau, N. R. (1993) *J Virol* **67**, 7229-7237
85. Heinzinger, N. K., Bukinsky, M. I., Haggerty, S. A., Ragland, A. M., Kewalramani, V., Lee, M. A., Gendelman, H. E., Ratner, L., Stevenson, M., and Emerman, M. (1994) *Proc Natl Acad Sci U S A* **91**, 7311-7315
86. Bartz, S. R., Rogel, M. E., and Emerman, M. (1996) *J Virol* **70**, 2324-2331
87. Ganser-Pornillos, B. K., Yeager, M., and Sundquist, W. I. (2008) *Curr Opin Struct Biol* **18**, 203-217
88. Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A., and Axel, R. (1986) *Cell* **47**, 333-348
89. Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M., and Berger, E. A. (1996) *Science* **272**, 1955-1958
90. Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P., and Paxton, W. A. (1996) *Nature* **381**, 667-673
91. Feng, Y., Broder, C. C., Kennedy, P. E., and Berger, E. A. (1996) *Science* **272**, 872-877

92. Huang, Y., Paxton, W. A., Wolinsky, S. M., Neumann, A. U., Zhang, L., He, T., Kang, S., Ceradini, D., Jin, Z., Yazdanbakhsh, K., Kunstman, K., Erickson, D., Dragon, E., Landau, N. R., Phair, J., Ho, D. D., and Koup, R. A. (1996) *Nat Med* **2**, 1240-1243
93. Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P. D., Wu, L., Mackay, C. R., LaRosa, G., Newman, W., Gerard, N., Gerard, C., and Sodroski, J. (1996) *Cell* **85**, 1135-1148
94. Hwang, S. S., Boyle, T. J., Lyerly, H. K., and Cullen, B. R. (1991) *Science* **253**, 71-74
95. Hartley, O., Klasse, P. J., Sattentau, Q. J., and Moore, J. P. (2005) *AIDS Res Hum Retroviruses* **21**, 171-189
96. Alkhatib, G., and Berger, E. A. (2007) *Eur J Med Res* **12**, 375-384
97. Santoro, F., Vassena, L., and Lusso, P. (2004) *New Microbiol* **27**, 17-29
98. Wu, L., Gerard, N. P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A. A., Desjardin, E., Newman, W., Gerard, C., and Sodroski, J. (1996) *Nature* **384**, 179-183
99. Jonckheere, H., Anne, J., and De Clercq, E. (2000) *Med Res Rev* **20**, 129-154
100. Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997) *Reverse Transcriptase and the Generation of Retroviral DNA in Retroviruses*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
101. Farnet, C. M., and Haseltine, W. A. (1991) *J Virol* **65**, 1910-1915
102. Stevens, M., De Clercq, E., and Balzarini, J. (2006) *Med Res Rev* **26**, 595-625
103. Cullen, B. R. (1998) *Cell* **93**, 685-692
104. Vastrik, I., D'Eustachio, P., Schmidt, E., Joshi-Tope, G., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., Matthews, L., Wu, G., Birney, E., and Stein, L. (2007) *Genome Biol* **8**, R39
105. Darlix, J. L., Lapadat-Tapolsky, M., de Rocquigny, H., and Roques, B. P. (1995) *J Mol Biol* **254**, 523-537
106. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science* **279**, 384-388
107. Huang, M., and Martin, M. A. (1997) *J Virol* **71**, 4472-4478
108. Ansari-Lari, M. A., and Gibbs, R. A. (1996) *J Virol* **70**, 3870-3875
109. Sandefur, S., Smith, R. M., Varthakavi, V., and Spearman, P. (2000) *J Virol* **74**, 7238-7249
110. Freed, E. O. (2001) *Somat Cell Mol Genet* **26**, 13-33
111. Vogt, V. M. (1996) *Curr Top Microbiol Immunol* **214**, 95-131
112. Kaplan, A. H., Manchester, M., and Swanstrom, R. (1994) *J Virol* **68**, 6782-6786
113. Darlix, J. L., Garrido, J. L., Morellet, N., Mely, Y., and de Rocquigny, H. (2007) *Adv Pharmacol* **55**, 299-346
114. Davis, J., Scherer, M., Tsai, W. P., and Long, C. (1976) *J Virol* **18**, 709-718
115. Darlix, J. L., and Spahr, P. F. (1982) *J Mol Biol* **160**, 147-161
116. Davis, N. L., and Rueckert, R. R. (1972) *J Virol* **10**, 1010-1020
117. Chen, M., Garon, C. F., and Papas, T. S. (1980) *Proc Natl Acad Sci U S A* **77**, 1296-1300
118. Meric, C., Gouilloud, E., and Spahr, P. F. (1988) *J Virol* **62**, 3328-3333
119. Gorelick, R. J., Henderson, L. E., Hanser, J. P., and Rein, A. (1988) *Proc Natl Acad Sci U S A* **85**, 8420-8424
120. Rein, A., Henderson, L. E., and Levin, J. G. (1998) *Trends Biochem Sci* **23**, 297-301

121. Jean-Luc Darlix, M. L. L., Yves Mély, Bernard Roques. (2002) Nucleocapsid Protein Chaperoning of Nucleic Acids at the Heart of HIV Structure, Assembly and cDNA Synthesis.
122. Gorelick, R. J., Gagliardi, T. D., Bosche, W. J., Wiltrout, T. A., Coren, L. V., Chabot, D. J., Lifson, J. D., Henderson, L. E., and Arthur, L. O. (1999) *Virology* **256**, 92-104
123. Berg, J. M. (1986) *Science* **232**, 485-487
124. Ottmann, M., Gabus, C., and Darlix, J. L. (1995) *J Virol* **69**, 1778-1784
125. Summers, M. F., South, T. L., Kim, B., and Hare, D. R. (1990) *Biochemistry* **29**, 329-340
126. Turner, B. G., and Summers, M. F. (1999) *J Mol Biol* **285**, 1-32
127. Thomas, J. A., and Gorelick, R. J. (2008) *Virus Res* **134**, 39-63
128. Leitner, T., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J.,Wolinksy, S., Korber, B. (2007) *Theoretical Biology and Biophysics Group, Los Alamos National Laboratory,* **LA-UR 07-7826, Los Alamos, NM**
129. Morellet, N., Jullian, N., De Rocquigny, H., Maigret, B., Darlix, J. L., and Roques, B. P. (1992) *EMBO J* **11**, 3059-3065
130. Gorelick, R. J., Chabot, D. J., Rein, A., Henderson, L. E., and Arthur, L. O. (1993) *J Virol* **67**, 4027-4036
131. Morellet, N., de Rocquigny, H., Mely, Y., Jullian, N., Demene, H., Ottmann, M., Gerard, D., Darlix, J. L., Fournie-Zaluski, M. C., and Roques, B. P. (1994) *J Mol Biol* **235**, 287-301
132. Adamson, C. S., and Freed, E. O. (2007) *Adv Pharmacol* **55**, 347-387
133. Lee, B. M., De Guzman, R. N., Turner, B. G., Tjandra, N., and Summers, M. F. (1998) *J Mol Biol* **279**, 633-649
134. Bombarda, E., Cherradi, H., Morellet, N., Roques, B. P., and Mely, Y. (2002) *Biochemistry* **41**, 4312-4320
135. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., and Summers, M. F. (2000) *J Mol Biol* **299**, 145-156
136. De Guzman, R. N., Turner, R. B., and Summers, M. F. (1998) *Biopolymers* **48**, 181-195
137. Cristofari, G., and Darlix, J. L. (2002) *Prog Nucleic Acid Res Mol Biol* **72**, 223-268
138. Levin, J. G., Guo, J., Rouzina, I., and Musier-Forsyth, K. (2005) *Prog Nucleic Acid Res Mol Biol* **80**, 217-286
139. Chan, B., Weidemaier, K., Yip, W. T., Barbara, P. F., and Musier-Forsyth, K. (1999) *Proc Natl Acad Sci U S A* **96**, 459-464
140. Erickson-Viitanen, S., Manfredi, J., Viitanen, P., Tribe, D. E., Tritch, R., Hutchison, C. A., 3rd, Loeb, D. D., and Swanstrom, R. (1989) *AIDS Res Hum Retroviruses* **5**, 577-591
141. Wondrak, E. M., Louis, J. M., de Rocquigny, H., Chermann, J. C., and Roques, B. P. (1993) *FEBS Lett* **333**, 21-24
142. Kafaie, J., Dolatshahi, M., Ajamian, L., Song, R., Mouland, A. J., Rouiller, I., and Laughrea, M. (2009) *Virology* **385**, 233-244
143. Chen, N., Morag, A., Almog, N., Blumenzweig, I., Dreazin, O., and Kotler, M. (2001) *J Gen Virol* **82**, 581-590
144. Pettit, S. C., Gulnik, S., Everitt, L., and Kaplan, A. H. (2003) *J Virol* **77**, 366-374
145. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., and Summers, M. F. (2000) *J Mol Biol* **301**, 491-511
146. Feng, Y. X., Campbell, S., Harvin, D., Ehresmann, B., Ehresmann, C., and Rein, A. (1999) *J Virol* **73**, 4251-4256

147. Bampi, C., Jacquenet, S., Lener, D., Decimo, D., and Darlix, J. L. (2004) *Int J Biochem Cell Biol* **36**, 1668-1686
148. Wacharapornin, P., Lauhakirti, D., and Auewarakul, P. (2007) *Virology* **358**, 48-54
149. Trono, D. (1992) *J Virol* **66**, 4893-4900
150. Lori, F., di Marzo Veronese, F., de Vico, A. L., Lusso, P., Reitz, M. S., Jr., and Gallo, R. C. (1992) *J Virol* **66**, 5067-5074
151. Barat, C., Schatz, O., Le Grice, S., and Darlix, J. L. (1993) *J Mol Biol* **231**, 185-190
152. Beerens, N., and Berkhout, B. (2000) *J Biol Chem* **275**, 15474-15481
153. Song, R., Kafaie, J., and Laughrea, M. (2008) *Biochemistry* **47**, 3283-3293
154. Peliska, J. A., Balasubramanian, S., Giedroc, D. P., and Benkovic, S. J. (1994) *Biochemistry* **33**, 13817-13823
155. Marzio, G., Tyagi, M., Gutierrez, M. I., and Giacca, M. (1998) *Proc Natl Acad Sci U S A* **95**, 13519-13524
156. Tanchou, V., Gabus, C., Rogemond, V., and Darlix, J. L. (1995) *J Mol Biol* **252**, 563-571
157. Guo, J., Wu, T., Anderson, J., Kane, B. F., Johnson, D. G., Gorelick, R. J., Henderson, L. E., and Levin, J. G. (2000) *J Virol* **74**, 8980-8988
158. Beltz, H., Azoulay, J., Bernacchi, S., Clamme, J. P., Ficheux, D., Roques, B., Darlix, J. L., and Mely, Y. (2003) *J Mol Biol* **328**, 95-108
159. Bernacchi, S., Stoylov, S., Piemont, E., Ficheux, D., Roques, B. P., Darlix, J. L., and Mely, Y. (2002) *J Mol Biol* **317**, 385-399
160. Basu, V. P., Song, M., Gao, L., Rigby, S. T., Hanson, M. N., and Bambara, R. A. (2008) *Virus Res* **134**, 19-38
161. Fisher, T. S., Darden, T., and Prasad, V. R. (2003) *J Mol Biol* **325**, 443-459
162. Bampi, C., Bibillo, A., Wendeler, M., Divita, G., Gorelick, R. J., Le Grice, S. F., and Darlix, J. L. (2006) *J Biol Chem* **281**, 11736-11743
163. Mikkelsen, J. G., and Pedersen, F. S. (2000) *J Biomed Sci* **7**, 77-99
164. Darlix, J. L., Vincent, A., Gabus, C., de Rocquigny, H., and Roques, B. (1993) *C R Acad Sci III* **316**, 763-771
165. Golinelli, M. P., and Hughes, S. H. (2002) *Virology* **294**, 122-134
166. Lener, D., Tanchou, V., Roques, B. P., Le Grice, S. F., and Darlix, J. L. (1998) *J Biol Chem* **273**, 33781-33786
167. Negroni, M., and Buc, H. (2000) *Proc Natl Acad Sci U S A* **97**, 6385-6390
168. DeStefano, J. J., Bambara, R. A., and Fay, P. J. (1994) *J Biol Chem* **269**, 161-168
169. Krishnamoorthy, G., Roques, B., Darlix, J. L., and Mely, Y. (2003) *Nucleic Acids Res* **31**, 5425-5432
170. Tanchou, V., Decimo, D., Pechoux, C., Lener, D., Rogemond, V., Berthoux, L., Ottmann, M., and Darlix, J. L. (1998) *J Virol* **72**, 4442-4447
171. Gorelick, R. J., Benveniste, R. E., Gagliardi, T. D., Wiltrout, T. A., Busch, L. K., Bosche, W. J., Coren, L. V., Lifson, J. D., Bradley, P. J., Henderson, L. E., and Arthur, L. O. (1999) *Virology* **253**, 259-270
172. Buckman, J. S., Bosche, W. J., and Gorelick, R. J. (2003) *J Virol* **77**, 1469-1480
173. Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L., and Charneau, P. (2000) *Cell* **101**, 173-185
174. Gorelick, R. J., Chabot, D. J., Ott, D. E., Gagliardi, T. D., Rein, A., Henderson, L. E., and Arthur, L. O. (1996) *J Virol* **70**, 2593-2597

175. Mirambeau, G., Lyonnais, S., Coulaud, D., Hameau, L., Lafosse, S., Jeusset, J., Borde, I., Reboud-Ravaux, M., Restle, T., Gorelick, R. J., and Le Cam, E. (2007) *PLoS ONE* **2**, e669
176. Gao, K., Gorelick, R. J., Johnson, D. G., and Bushman, F. (2003) *J Virol* **77**, 1598-1603
177. Le Cam, E., Coulaud, D., Delain, E., Petitjean, P., Roques, B. P., Gerard, D., Stoylova, E., Vuilleumier, C., Stoylov, S. P., and Mely, Y. (1998) *Biopolymers* **45**, 217-229
178. Gabus, C., Auxilien, S., Pechoux, C., Dormont, D., Swietnicki, W., Morillas, M., Surewicz, W., Nandi, P., and Darlix, J. L. (2001) *J Mol Biol* **307**, 1011-1021
179. Saliou, J. M., Bourgeois, C. F., Ayadi-Ben Mena, L., Ropers, D., Jacquenet, S., Marchand, V., Stevenin, J., and Branlant, C. (2009) *Front Biosci* **14**, 2714-2729
180. Zhang, J., and Crumpacker, C. S. (2002) *J Virol* **76**, 10444-10454
181. Cullen, B. R. (2005) *Nature* **433**, 26-27
182. Blot, G., Janvier, K., Le Panse, S., Benarous, R., and Berlioz-Torrent, C. (2003) *J Virol* **77**, 6931-6945
183. Joshi, A., Ablan, S. D., Soheilian, F., Nagashima, K., and Freed, E. O. (2009) *J Virol*
184. Briggs, J. A., Simon, M. N., Gross, I., Krausslich, H. G., Fuller, S. D., Vogt, V. M., and Johnson, M. C. (2004) *Nat Struct Mol Biol* **11**, 672-675
185. Brasey, A., Lopez-Lastra, M., Ohlmann, T., Beerens, N., Berkhout, B., Darlix, J. L., and Sonenberg, N. (2003) *J Virol* **77**, 3939-3949
186. Goh, W. C., Rogel, M. E., Kinsey, C. M., Michael, S. F., Fultz, P. N., Nowak, M. A., Hahn, B. H., and Emerman, M. (1998) *Nat Med* **4**, 65-71
187. Herbreteau, C. H., Weill, L., Decimo, D., Prevot, D., Darlix, J. L., Sargueil, B., and Ohlmann, T. (2005) *Nat Struct Mol Biol* **12**, 1001-1007
188. Freed, E. O. (2002) *J Virol* **76**, 4679-4687
189. Gheysen, D., Jacobs, E., de Foresta, F., Thiriart, C., Francotte, M., Thines, D., and De Wilde, M. (1989) *Cell* **59**, 103-112
190. Accola, M. A., Strack, B., and Gottlinger, H. G. (2000) *J Virol* **74**, 5395-5402
191. Borsetti, A., Ohagen, A., and Gottlinger, H. G. (1998) *J Virol* **72**, 9313-9317
192. Cimarelli, A., Sandin, S., Hoglund, S., and Luban, J. (2000) *J Virol* **74**, 3046-3057
193. Burniston, M. T., Cimarelli, A., Colgan, J., Curtis, S. P., and Luban, J. (1999) *J Virol* **73**, 8527-8540
194. Accola, M. A., Hoglund, S., and Gottlinger, H. G. (1998) *J Virol* **72**, 2072-2078
195. Mammano, F., Ohagen, A., Hoglund, S., and Gottlinger, H. G. (1994) *J Virol* **68**, 4927-4936
196. Hatziioannou, T., Perez-Caballero, D., Cowan, S., and Bieniasz, P. D. (2005) *J Virol* **79**, 176-183
197. Muriaux, D., Mirro, J., Harvin, D., and Rein, A. (2001) *Proc Natl Acad Sci U S A* **98**, 5246-5251
198. Demirov, D. G., Orenstein, J. M., and Freed, E. O. (2002) *J Virol* **76**, 105-117
199. Campbell, S., Fisher, R. J., Towler, E. M., Fox, S., Issaq, H. J., Wolfe, T., Phillips, L. R., and Rein, A. (2001) *Proc Natl Acad Sci U S A* **98**, 10875-10879
200. Sandefur, S., Varthakavi, V., and Spearman, P. (1998) *J Virol* **72**, 2723-2732
201. Ott, D. E., Coren, L. V., Chertova, E. N., Gagliardi, T. D., Nagashima, K., Sowder, R. C., 2nd, Poon, D. T., and Gorelick, R. J. (2003) *J Virol* **77**, 5547-5556
202. Berkhout, B. (1996) *Prog Nucleic Acid Res Mol Biol* **54**, 1-34
203. Hayashi, T., Shioda, T., Iwakura, Y., and Shibuta, H. (1992) *Virology* **188**, 590-599

204. Cruceanu, M., Urbaneja, M. A., Hixson, C. V., Johnson, D. G., Datta, S. A., Fivash, M. J., Stephen, A. G., Fisher, R. J., Gorelick, R. J., Casas-Finet, J. R., Rein, A., Rouzina, I., and Williams, M. C. (2006) *Nucleic Acids Res* **34**, 593-605
205. Huthoff, H., and Berkhout, B. (2001) *RNA* **7**, 143-157
206. Muriaux, D., Mirro, J., Nagashima, K., Harvin, D., and Rein, A. (2002) *J Virol* **76**, 11405-11413
207. Coffin, J. M., Hughes, S. H., and Varmus, H. E. . (1997) *Retroviruses. Synthesis and processing of viral RNA*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
208. Ott, D. E., Coren, L. V., and Gagliardi, T. D. (2005) *J Virol* **79**, 13839-13847
209. Poon, D. T., Li, G., and Aldovini, A. (1998) *J Virol* **72**, 1983-1993
210. D'Souza, V., and Summers, M. F. (2005) *Nat Rev Microbiol* **3**, 643-655
211. Berkowitz, R., Fisher, J., and Goff, S. P. (1996) *Curr Top Microbiol Immunol* **214**, 177-218
212. Anderson, E. C., and Lever, A. M. (2006) *J Virol* **80**, 10478-10486
213. Poon, D. T., Chertova, E. N., and Ott, D. E. (2002) *Virology* **293**, 368-378
214. Harrison, G. P., Miele, G., Hunter, E., and Lever, A. M. (1998) *J Virol* **72**, 5886-5896
215. Rulli, S. J., Jr., Hibbert, C. S., Mirro, J., Pederson, T., Biswal, S., and Rein, A. (2007) *J Virol* **81**, 6623-6631
216. Amarasinghe, G. K., Zhou, J., Miskimon, M., Chancellor, K. J., McDonald, J. A., Matthews, A. G., Miller, R. R., Rouse, M. D., and Summers, M. F. (2001) *J Mol Biol* **314**, 961-970
217. Nguyen, D. G., Booth, A., Gould, S. J., and Hildreth, J. E. (2003) *J Biol Chem* **278**, 52347-52354
218. Freed, E. O., and Mouland, A. J. (2006) *Retrovirology* **3**, 77
219. McLaughlin, S., and Murray, D. (2005) *Nature* **438**, 605-611
220. Hill, M., Tachedjian, G., and Mak, J. (2005) *Curr HIV Res* **3**, 73-85
221. Cen, S., Niu, M., Saadatmand, J., Guo, F., Huang, Y., Nabel, G. J., and Kleiman, L. (2004) *J Virol* **78**, 1042-1049
222. Khorchid, A., Halwani, R., Wainberg, M. A., and Kleiman, L. (2002) *J Virol* **76**, 4131-4137
223. Sakuragi, J., Ueda, S., Iwamoto, A., and Shioda, T. (2003) *J Virol* **77**, 4060-4069
224. Bukrinskaya, A. G. (2004) *Arch Virol* **149**, 1067-1082
225. Marquet, R., Isel, C., Ehresmann, C., and Ehresmann, B. (1995) *Biochimie* **77**, 113-124
226. Cen, S., Javanbakht, H., Kim, S., Shiba, K., Craven, R., Rein, A., Ewalt, K., Schimmel, P., Musier-Forsyth, K., and Kleiman, L. (2002) *J Virol* **76**, 13111-13115
227. Kleiman, L., and Cen, S. (2004) *Int J Biochem Cell Biol* **36**, 1776-1786
228. Cen, S., Huang, Y., Khorchid, A., Darlix, J. L., Wainberg, M. A., and Kleiman, L. (1999) *J Virol* **73**, 4485-4488
229. Lu, Y. L., Bennett, R. P., Wills, J. W., Gorelick, R., and Ratner, L. (1995) *J Virol* **69**, 6873-6879
230. Andersen, J. L., and Planelles, V. (2005) *Curr HIV Res* **3**, 43-51
231. Chiu, Y. L., and Greene, W. C. (2006) *J Biol Chem* **281**, 8309-8312
232. Luo, K., Liu, B., Xiao, Z., Yu, Y., Yu, X., Gorelick, R., and Yu, X. F. (2004) *J Virol* **78**, 11841-11852
233. Henriet, S., Richer, D., Bernacchi, S., Decroly, E., Vigne, R., Ehresmann, B., Ehresmann, C., Paillart, J. C., and Marquet, R. (2005) *J Mol Biol* **354**, 55-72

234. Holm, K., Weclewicz, K., Hewson, R., and Suomalainen, M. (2003) *J Virol* **77**, 4805-4817
235. Mammano, F., Kondo, E., Sodroski, J., Bukovsky, A., and Gottlinger, H. G. (1995) *J Virol* **69**, 3824-3830
236. Garnier, L., Bowzard, J. B., and Wills, J. W. (1998) *AIDS* **12 Suppl A**, S5-16
237. Garoff, H., Hewson, R., and Opstelten, D. J. (1998) *Microbiol Mol Biol Rev* **62**, 1171-1190
238. Sheng, N., Pettit, S. C., Tritch, R. J., Ozturk, D. H., Rayner, M. M., Swanstrom, R., and Erickson-Viitanen, S. (1997) *J Virol* **71**, 5723-5732
239. Poon, D. T., Wu, J., and Aldovini, A. (1996) *J Virol* **70**, 6607-6616
240. Liang, C., Rong, L., Cherry, E., Kleiman, L., Laughrea, M., and Wainberg, M. A. (1999) *J Virol* **73**, 6147-6151
241. Butsch, M., and Boris-Lawrie, K. (2002) *J Virol* **76**, 3089-3094
242. Buck, C. B., Shen, X., Egan, M. A., Pierson, T. C., Walker, C. M., and Siliciano, R. F. (2001) *J Virol* **75**, 181-191
243. Abbink, T. E., and Berkhout, B. (2003) *J Biol Chem* **278**, 11601-11611
244. Baba, S., Takahashi, K., Koyanagi, Y., Yamamoto, N., Takaku, H., Gorelick, R. J., and Kawai, G. (2003) *J Biochem* **134**, 637-639
245. Takahashi, K., Baba, S., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2001) *J Biol Chem* **276**, 31274-31278
246. Laughrea, M., Jette, L., Mak, J., Kleiman, L., Liang, C., and Wainberg, M. A. (1997) *J Virol* **71**, 3397-3406
247. Chin, M. P., Chen, J., Nikolaitchik, O. A., and Hu, W. S. (2007) *Virology* **363**, 437-446
248. Chin, M. P., Rhodes, T. D., Chen, J., Fu, W., and Hu, W. S. (2005) *Proc Natl Acad Sci U S A* **102**, 9002-9007
249. Russell, R. S., Liang, C., and Wainberg, M. A. (2004) *Retrovirology* **1**, 23
250. Meric, C., and Spahr, P. F. (1986) *J Virol* **60**, 450-459
251. Muriaux, D., De Rocquigny, H., Roques, B. P., and Paoletti, J. (1996) *J Biol Chem* **271**, 33686-33692
252. Laughrea, M., Shen, N., Jette, L., Darlix, J. L., Kleiman, L., and Wainberg, M. A. (2001) *Virology* **281**, 109-116
253. Song, R., Kafaie, J., Yang, L., and Laughrea, M. (2007) *J Mol Biol* **371**, 1084-1098
254. Bowles, N. E., Damay, P., and Spahr, P. F. (1993) *J Virol* **67**, 623-631
255. Dupraz, P., Oertle, S., Meric, C., Damay, P., and Spahr, P. F. (1990) *J Virol* **64**, 4978-4987
256. Fu, W., and Rein, A. (1993) *J Virol* **67**, 5443-5449
257. Shen, N., Jette, L., Wainberg, M. A., and Laughrea, M. (2001) *J Virol* **75**, 10543-10549
258. Clever, J. L., and Parslow, T. G. (1997) *J Virol* **71**, 3407-3414
259. Haddrick, M., Lear, A. L., Cann, A. J., and Heaphy, S. (1996) *J Mol Biol* **259**, 58-68
260. Shen, N., Jette, L., Liang, C., Wainberg, M. A., and Laughrea, M. (2000) *J Virol* **74**, 5729-5735
261. Hagan, N. A., and Fabris, D. (2007) *J Mol Biol* **365**, 396-410
262. Mihailescu, M. R., and Marino, J. P. (2004) *Proc Natl Acad Sci U S A* **101**, 1189-1194
263. Mujeeb, A., Ulyanov, N. B., Georgantis, S., Smirnov, I., Chung, J., Parslow, T. G., and James, T. L. (2007) *Nucleic Acids Res* **35**, 2026-2034
264. Rist, M. J., and Marino, J. P. (2002) *Biochemistry* **41**, 14762-14770

265. Takahashi, K. I., Baba, S., Chattopadhyay, P., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2000) *RNA* **6**, 96-102
266. Turner, K. B., Hagan, N. A., and Fabris, D. (2007) *J Mol Biol* **369**, 812-828
267. Berg, J. M., and Shi, Y. (1996) *Science* **271**, 1081-1085
268. Bess, J. W., Jr., Powell, P. J., Issaq, H. J., Schumack, L. J., Grimes, M. K., Henderson, L. E., and Arthur, L. O. (1992) *J Virol* **66**, 840-847
269. Green, L. M., and Berg, J. M. (1989) *Proc Natl Acad Sci U S A* **86**, 4047-4051
270. Laughrea, M., and Jette, L. (1994) *Biochemistry* **33**, 13464-13474
271. Cimarelli, A., and Luban, J. (2001) *J Virol* **75**, 7193-7197
272. Krogstad, P., Geng, Y. Z., Rey, O., Canon, J., Ibarrondo, F. J., Ackerson, B., Patel, J., and Aldovini, A. (2002) *Virology* **294**, 282-288
273. Adachi, A., Gendelman, H. E., Koenig, S., Folks, T., Willey, R., Rabson, A., and Martin, M. A. (1986) *J Virol* **59**, 284-291
274. Aldovini, A., and Young, R. A. (1990) *J Virol* **64**, 1920-1926
275. Demene, H., Dong, C. Z., Ottmann, M., Rouyez, M. C., Jullian, N., Morellet, N., Mely, Y., Darlix, J. L., Fournie-Zaluski, M. C., Saragosti, S., and et al. (1994) *Biochemistry* **33**, 11707-11716
276. Dorfman, T., Luban, J., Goff, S. P., Haseltine, W. A., and Gottlinger, H. G. (1993) *J Virol* **67**, 6159-6169
277. Gorelick, R. J., Nigida, S. M., Jr., Bess, J. W., Jr., Arthur, L. O., Henderson, L. E., and Rein, A. (1990) *J Virol* **64**, 3207-3211
278. Dawson, L., and Yu, X. F. (1998) *Virology* **251**, 141-157
279. Wang, S. W., and Aldovini, A. (2002) *J Virol* **76**, 11853-11865
280. Wang, S. W., Noonan, K., and Aldovini, A. (2004) *J Virol* **78**, 716-723
281. Ramboarina, S., Morellet, N., Fournie-Zaluski, M. C., and Roques, B. P. (1999) *Biochemistry* **38**, 9600-9607
282. Mely, Y., De Rocquigny, H., Morellet, N., Roques, B. P., and Gerad, D. (1996) *Biochemistry* **35**, 5175-5182
283. Berg, J. M. (1990) *Annu Rev Biophys Biophys Chem* **19**, 405-421
284. Miller, J., McLachlan, A. D., and Klug, A. (1985) *EMBO J* **4**, 1609-1614
285. Pabo, C. O., Peisach, E., and Grant, R. A. (2001) *Annu Rev Biochem* **70**, 313-340
286. Buxton, P., Tachedjian, G., and Mak, J. (2005) *J Virol* **79**, 6338-6348
287. Guo, J., Wu, T., Kane, B. F., Johnson, D. G., Henderson, L. E., Gorelick, R. J., and Levin, J. G. (2002) *J Virol* **76**, 4370-4378
288. Williams, M. C., Gorelick, R. J., and Musier-Forsyth, K. (2002) *Proc Natl Acad Sci U S A* **99**, 8614-8619
289. Sakuragi, J., Shioda, T., and Panganiban, A. T. (2001) *J Virol* **75**, 2557-2565
290. Housset, V., De Rocquigny, H., Roques, B. P., and Darlix, J. L. (1993) *J Virol* **67**, 2537-2545
291. South, T. L., Blake, P. R., Hare, D. R., and Summers, M. F. (1991) *Biochemistry* **30**, 6342-6349
292. Meric, C., and Goff, S. P. (1989) *J Virol* **63**, 1558-1568
293. McGrath, C. F., Buckman, J. S., Gagliardi, T. D., Bosche, W. J., Coren, L. V., and Gorelick, R. J. (2003) *J Virol* **77**, 8524-8531
294. Leitner, T., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J. W., Wolinksky, , and S. & Korber, B. (2005) *Los Alamos National*

295. Fu, W., Dang, Q., Nagashima, K., Freed, E. O., Pathak, V. K., and Hu, W. S. (2006) *J Virol* **80**, 1242-1249
296. Hibbert, C. S., Mirro, J., and Rein, A. (2004) *J Virol* **78**, 10927-10938
297. Oshima, M., Muriaux, D., Mirro, J., Nagashima, K., Dryden, K., Yeager, M., and Rein, A. (2004) *J Virol* **78**, 1411-1420
298. Shehu-Xhilaga, M., Crowe, S. M., and Mak, J. (2001) *J Virol* **75**, 1834-1841
299. Shehu-Xhilaga, M., Hill, M., Marshall, J. A., Kappes, J., Crowe, S. M., and Mak, J. (2002) *J Virol* **76**, 4331-4340
300. Gorelick, R. J., Fu, W., Gagliardi, T. D., Bosche, W. J., Rein, A., Henderson, L. E., and Arthur, L. O. (1999) *J Virol* **73**, 8185-8195
301. Mizuno, A., Ido, E., Goto, T., Kuwata, T., Nakai, M., and Hayami, M. (1996) *AIDS Res Hum Retroviruses* **12**, 793-800
302. Schwartz, M. D., Fiore, D., and Panganiban, A. T. (1997) *J Virol* **71**, 9295-9305
303. Akahata, W., Ido, E., Akiyama, H., Uesaka, H., Enose, Y., Horiuchi, R., Kuwata, T., Goto, T., Takahashi, H., and Hayami, M. (2003) *J Gen Virol* **84**, 2237-2244
304. Yovandich, J. L., Chertova, E. N., Kane, B. P., Gagliardi, T. D., Bess, J. W., Jr., Sowder, R. C., 2nd, Henderson, L. E., and Gorelick, R. J. (2001) *J Virol* **75**, 115-124
305. Fu, W., Prasad, V. V., Chen, J., Nikolaitchik, O., and Hu, W. S. (2007) *Virology* **363**, 210-219
306. Dannull, J., Surovoy, A., Jung, G., and Moelling, K. (1994) *EMBO J* **13**, 1525-1533
307. Krizek, B. A., Zawadzke, L. E., and Berg, J. M. (1993) *Protein Sci* **2**, 1313-1319
308. Vallee, B. L., and Auld, D. S. (1990) *Biochemistry* **29**, 5647-5659
309. Stote, R. H., Kellenberger, E., Muller, H., Bombarda, E., Roques, B. P., Kieffer, B., and Mely, Y. (2004) *Biochemistry* **43**, 7687-7697
310. Morellet, N., Demene, H., Teilleux, V., Huynh-Dinh, T., de Rocquigny, H., Fournie-Zaluski, M. C., and Roques, B. P. (1998) *J Mol Biol* **283**, 419-434
311. Laughrea, M., and Jette, L. (1996) *Biochemistry* **35**, 1589-1598
312. Pettit, S. C., Everitt, L. E., Choudhury, S., Dunn, B. M., and Kaplan, A. H. (2004) *J Virol* **78**, 8477-8485
313. Mervis, R. J., Ahmad, N., Lillehoj, E. P., Raum, M. G., Salazar, F. H., Chan, H. W., and Venkatesan, S. (1988) *J Virol* **62**, 3993-4002
314. Gowda, S. D., Stein, B. S., Steimer, K. S., and Engleman, E. G. (1989) *J Virol* **63**, 1451-1454
315. Pettit, S. C., Henderson, G. J., Schiffer, C. A., and Swanstrom, R. (2002) *J Virol* **76**, 10226-10233
316. Croteau, G., Doyon, L., Thibeault, D., McKercher, G., Pilote, L., and Lamarre, D. (1997) *J Virol* **71**, 1089-1096
317. Rose, R. B., Craik, C. S., Douglas, N. L., and Stroud, R. M. (1996) *Biochemistry* **35**, 12933-12944
318. Schock, H. B., Garsky, V. M., and Kuo, L. C. (1996) *J Biol Chem* **271**, 31957-31963
319. Carrillo, A., Stewart, K. D., Sham, H. L., Norbeck, D. W., Kohlbrenner, W. E., Leonard, J. M., Kempf, D. J., and Molla, A. (1998) *J Virol* **72**, 7532-7541
320. Zhang, Y., and Barklis, E. (1997) *J Virol* **71**, 6765-6776
321. Doyon, L., Croteau, G., Thibeault, D., Poulin, F., Pilote, L., and Lamarre, D. (1996) *J Virol* **70**, 3763-3769

322. Mammano, F., Petit, C., and Clavel, F. (1998) *J Virol* **72**, 7632-7637
323. Chassagne, J., Verrelle, P., Dionet, C., Clavel, F., Barre-Sinoussi, F., Chermann, J. C., Montagnier, L., Gluckman, J. C., and Klatzmann, D. (1986) *J Immunol* **136**, 1442-1445
324. Veronese, F. D., Rahman, R., Copeland, T. D., Oroszlan, S., Gallo, R. C., and Sarngadharan, M. G. (1987) *AIDS Res Hum Retroviruses* **3**, 253-264
325. Henderson, L. E., Bowers, M. A., Sowder, R. C., 2nd, Serabyn, S. A., Johnson, D. G., Bess, J. W., Jr., Arthur, L. O., Bryant, D. K., and Fenselau, C. (1992) *J Virol* **66**, 1856-1865
326. Kaplan, A. H., and Swanstrom, R. (1991) *Proc Natl Acad Sci U S A* **88**, 4528-4532
327. Sheng, N., and Erickson-Viitanen, S. (1994) *J Virol* **68**, 6207-6214
328. Tozser, J., Blaha, I., Copeland, T. D., Wondrak, E. M., and Oroszlan, S. (1991) *FEBS Lett* **281**, 77-80
329. Dinman, J. D., Richter, S., Plant, E. P., Taylor, R. C., Hammell, A. B., and Rana, T. M. (2002) *Proc Natl Acad Sci U S A* **99**, 5331-5336
330. Dulude, D., Baril, M., and Brakier-Gingras, L. (2002) *Nucleic Acids Res* **30**, 5094-5102
331. Tessmer, U., and Krausslich, H. G. (1998) *J Virol* **72**, 3459-3463
332. Babe, L. M., Rose, J., and Craik, C. S. (1995) *Proc Natl Acad Sci U S A* **92**, 10069-10073
333. DiIanni, C. L., Davis, L. J., Holloway, M. K., Herber, W. K., Darke, P. L., Kohl, N. E., and Dixon, R. A. (1990) *J Biol Chem* **265**, 17348-17354
334. Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., and et al. (1989) *Nature* **342**, 299-302
335. Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) *Nature* **337**, 615-620
336. Pettit, S. C., Simsic, J., Loeb, D. D., Everitt, L., Hutchison, C. A., 3rd, and Swanstrom, R. (1991) *J Biol Chem* **266**, 14539-14547
337. Murphy, E. M., Jimenez, H. R., and Smith, S. M. (2008) *Adv Pharmacol* **56**, 27-73
338. Kaplan, A. H., Zack, J. A., Knigge, M., Paul, D. A., Kempf, D. J., Norbeck, D. W., and Swanstrom, R. (1993) *J Virol* **67**, 4050-4055
339. Konvalinka, J., Litterst, M. A., Welker, R., Kottler, H., Rippmann, F., Heuser, A. M., and Krausslich, H. G. (1995) *J Virol* **69**, 7180-7186
340. Moore, M. D., Fu, W., Soheilian, F., Nagashima, K., Ptak, R. G., Pathak, V. K., and Hu, W. S. (2008) *Virology* **379**, 152-160
341. Pettit, S. C., Clemente, J. C., Jeung, J. A., Dunn, B. M., and Kaplan, A. H. (2005) *J Virol* **79**, 10601-10607
342. Rose, J. R., Babe, L. M., and Craik, C. S. (1995) *J Virol* **69**, 2751-2758
343. Hargittai, M. R., Gorelick, R. J., Rouzina, I., and Musier-Forsyth, K. (2004) *J Mol Biol* **337**, 951-968
344. Hargittai, M. R., Mangla, A. T., Gorelick, R. J., and Musier-Forsyth, K. (2001) *J Mol Biol* **312**, 985-997
345. Rong, L., Liang, C., Hsu, M., Guo, X., Roques, B. P., and Wainberg, M. A. (2001) *J Biol Chem* **276**, 47725-47732
346. Darlix, J. L., Mély, Y. and Roques, B. (2002) http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2002/partI/Darlix.pdf.
347. Thomas, J. A., Gagliardi, T. D., Alvord, W. G., Lubomirski, M., Bosche, W. J., and Gorelick, R. J. (2006) *Virology* **353**, 41-51

348. Gross, I., Hohenberg, H., Wilk, T., Wiegers, K., Grattinger, M., Muller, B., Fuller, S., and Krausslich, H. G. (2000) *EMBO J* **19**, 103-113
349. Wiegers, K., Rutter, G., Kottler, H., Tessmer, U., Hohenberg, H., and Krausslich, H. G. (1998) *J Virol* **72**, 2846-2854
350. Wright, E. R., Schooler, J. B., Ding, H. J., Kieffer, C., Fillmore, C., Sundquist, W. I., and Jensen, G. J. (2007) *EMBO J* **26**, 2218-2226
351. Zabransky, A., Hunter, E., and Sakalian, M. (2002) *Virology* **294**, 141-150