2010

Dissecting the Contribution of the Carboxyl-Terminal Domain and Tail of HIV-1 Integrase to Viral Dynamics and Enzymatic Function

Kevin Dominic Mohammed

Dissecting the Contribution of the Carboxyl-Terminal Domain and Tail of HIV-1 Integrase to Viral Dynamics and Enzymatic Function

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Kevin Dominic Mohammed

June 2010

Dissecting the Contribution of the Carboxyl-Terminal Domain and Tail of HIV-1 Integrase to Viral Dynamics and Enzymatic Function

Kevin Dominic Mohammed, Ph.D.

The Rockefeller University 2010

The Human Immunodeficiency Virus Type-1 (HIV-1) is the causative agent of the Acquired Immune Deficiency Syndrome (AIDS). Combination antiviral therapy has proven to be particularly effective at suppressing viral replication, yet complete eradication of the virus from an infected individual remains elusive. Recently, a new class of antiviral drugs targeting the viral integrase (IN) has been added to the HAART (high active antiretroviral therapy) regimen. This novel drug class exerts its inhibitory effect by targeting one aspect of the dual-staged integration reaction. In contrast to the other two viral targets of HAART, the reverse transcriptase (RT) and protease (PR) enzymes, both of which have singular roles in viral replication, IN contributes a plurality of functions to the viral life cycle, thus potentiating the development for further therapeutic intervention. In spite of the intensity of investigation in the field, the mechanistic details of several aspects of IN activity, especially of its ancillary functions remain unclear. This thesis work focuses primarily on dissecting the contribution of the carboxyl-terminal domain (CTD), in particular the terminal 18 amino acid 'tail' of this domain, to both the primary integration activities and secondary functions of IN. This

was accomplished by conducting a survey of incremental deletions aimed at gradually removing this region of the protein, and testing the resulting viruses in a number of measurable assays of virologic function. In so doing we identified a number of anomalous mutant phenotypes, the ensuing characterization of which should contribute to better understanding the roles of IN with regard to both its primary and secondary functions in the replicative strategy of HIV-1.

For my loving family, my mother Janet, and my siblings Hamish and Tracey, and for my father, the best chemistry teacher to have lived

ACKNOWLEDGMENTS

I am deeply grateful to have been given the opportunity to work with Dr. Mark Muesing throughout my graduate career. His mentorship, guidance and friendship have been instrumental in my development in the past six years, and his imagination and passion for science is a source an inspiration for any young researcher.

To members Drs. Charlie Rice, Paul Bieniasz, and Magda Konarska of my thesis advisory committee I am indebted for their sound advice and support over the years. It has been my privilege to have had them offer their criticisms and encouragement. I would like to extend my gratitude to Dr. Monica Roth, who agreed to evaluate my thesis as an external committee member, for taking the time out of her schedule to lend her expertise for my defense.

I thank all the past and current members of the Muesing lab, especially Michael Topper, whose technical expertise, attention to detail and instruction helped me tremendously when it came time for me to start thinking independently about science and for conducting and criticizing my own experiments. To the friends and colleagues I've made at the Aaron Diamond AIDS Research Center and at the Rockefeller University, and to those with whom I started this chapter of my life, I thank immeasurably for all their help, support and friendship, which all made for my time here in New York and at school such a rewarding experience.

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

ABBREVIATIONS

3’ P - 3’ Processing

3’ PPT- 3’ Polypurine Tract

AIDS - Acquired Immunodeficiency Syndrome

bp - base pairs

Bsd^r - Blastocidin D Resistance

bsd - Blastocidin D Resistance Gene

CA - Capsid

cDNA - complementary DNA

cPPT – Central Polypurine Tract

CCD - Catalytic Core Domain

CTD – Carboxyl Terminal Domain

DMEM - Dulbecco's Modified Eagle's Medium

EGFP – Enhanced Green Fluorescent Protein

Env- Envelope

FBS – Fetal Bovine Serum

GFP – Green Fluorescent Protein

GHOST – LTR-GFP Reporter Containing Human Osteosarcoma Cell Line

HEK – Human Embryonic Kidney

HIV-1 – Human Immunodeficiency Virus Type 1

His – Histidine

HTLV-1, II - Human T-cell Leukemia Virus

IN – Integrase

INI1 – Integrase Interactor 1

LAV - Lymphadenopathy Virus

LEDGF/p75 – Lens Epithelium-Derived Growth Factor p75

LTR – Long Terminal Repeat

MA - Matrix

MBP – Maltose Binding Protein

MSD- Major Splice Donor

NNRTI - Non-Nucleoside RT Inhibitor

NPC – Nuclear Pore Complex

NTD – Amino-Terminal Domain

ORF- Open Reading Frame

PBMC- Peripheral Blood Mononuclear Cells

PBS – Phosphate Buffered Saline

PCR – Polymerase Chain Reaction

PEI – Polyethylenimine

Pen/Strep – Penicillin/Streptomycin

PVDF – Polyvinylidene Difluoride

R – Repeat Sequence

Rev Response Element

RT – Reverse Transcription or Transcriptase

SDS – Sodium Dodecyl Sulfate

SH3 – Src Homology-3

TAR – Transactivation Response Element

U3 – Unique 3’ Sequence

U5 – Unique 5’ Sequence

Vif – Viral Infectivity Factor

VLP – Virus-Like Particle

Vpr – Viral Protein R

Vpu – Viral Protein U

VSV-G – Vesicular Stomatitis Virus G Protein

CHAPTER 1- INTRODUCTION

Despite the recent incidence of HIV-1 as a human pathogen, it has, in the ensuing two and a half decades since its inception, become one the most prominent public health challenges in modern times, with current estimates of 30 million infected people worldwide [1]. Without treatment, HIV-1 infection in the overwhelming majority of cases progresses ultimately to AIDS, the end stage of disease in which the host's substantially ablated immune system renders the individual susceptible to a wide range of opportunistic infections. Antiviral drugs that target and effectively inhibit viral replication have proven to significantly curtail the disease progression and concomitantly decrease HIV-1-associated mortality, but efforts to completely eradicate the virus from the body has thus far been unsuccessful.

The IN protein has long been a much sought-after target for antiviral therapy. The reaction that it catalyses, integration, has no counterpart to any known cellular processes, thus increasing the likelihood of developing highly specific, well-tolerated drugs which would block this viral activity. The development in recent years of compounds that selectively target IN activity has successfully increased the repertoire of available antiviral drugs, and proven an extremely valuable addition to the existing armamentarium particularly to patients who are unresponsive to previous drug classes. These compounds, chemically classified as diketo acids, inhibit viral replication by targeting the second step

of the integration reaction, called strand transfer, and thus the designation of these drugs as strand transfer inhibitors (STI).

The impetus to develop new drugs is ever present, however, as the genetic barrier to drug resistance is low, requiring only one or two point mutations to confer resistance to both developed IN inhibitors. In light of the highly specific targeting of the STIs, IN is very likely amenable to inhibition by other routes. Indeed the initiation of strand transfer is entirely dependent upon the completion of a prior catalytic step (3' processing), thus making the latter an equally attractive enzymatic target for directed IN inhibition.

In addition to its catalytic roles in proviral integration and establishment, IN facilitates a number of secondary functions during the viral life cycle that include viral assembly, the precision of particle formation, and the efficiency of reverse transcription [2-8], each of which could present a future target for therapeutic intervention. Though significant insights have been gained in the mechanistic details of the integration reaction, relatively little is known of IN involvement in its secondary activities.

Efforts to elicit a high-resolution crystal structure of HIV-1 IN have thus far been unsuccessful, but evidence gleaned from other lines of investigation indicate that IN consists of three domains: an N-terminal domain (NTD), a catalytic core domain (CCD) and a carboxyl-terminal domain (CTD). Each domain contributes in some fashion to the mechanics of the integration itself, but it is apparent that the CTD is especially involved in the facilitation of secondary functions. Indeed, mutagenic surveys of conserved residues in the CTD revealed that the vast majority of IN mutants tested were defective for viral functions that occur prior to integration, at or before reverse transcription or

during the late stages of viral polyprotein processing and/or virion maturation and egress [2].

It was recently discovered that the IN CTD is a potent target for p300-mediated histone acetyltransferase (HAT) activity, targeting a cluster three lysine residues (Lys 264, 266, 273) for acetylation [3, 4]. Though we demonstrated that this post-translational modification was not critical for viral propagation in a tissue culture infection model [4], we endeavor to investigate this phenomenon further. The final 18 amino acids of the IN CTD constitute a region of the domain that is recalcitrant to structural determination, due its high levels of flexibility and molecular motion. This unstructured 'tail' encompasses residues 271-288, and is reminiscent of the unstructured tails of the histone proteins which themselves undergo extensive covalent modifications that modulate their various activities [5]. In light of the multifunctional nature of IN, and the potential role of the CTD to IN secondary functions, it was intriguing to find that the CTD is such a potent target for covalent modifications, alterations that could feasibly modulate the enzyme's activity. To investigate this phenomenon, we created a sequential panel of truncated IN mutants, with the aim of gradually removing the unstructured tail, and consequently the residues targeted for modification, and observing the effect on a variety of measurable viral processes.

We found that sequential and stepwise removal of residues resulted in a gradual attenuation of infectivity up to residue 273 (IN 273), corresponding to a removal of fifteen residues, and further shortening of the CTD resulted in a sudden loss of infectivity (IN 272). Removal of the two successive residues, corresponding to truncation at IN 270

and 271, still rendered these viruses noninfectious. However and remarkably, it was discovered that removal the one more consecutive residue (D270), corresponding to a truncation that includes IN R269, rescued viral infectivity. This thesis reports the investigation of IN C-terminal truncations within a window of 28 residues (IN 260-288), wherein these mutants are characterized and classified via a number of assays in order to ascertain the contributions that the IN CTD makes to IN function, and in a broader sense to HIV-1 replication itself. This information is not only of significant academic interest, but of practical importance as well, for the potential development of combative agents that more comprehensively inhibit one or more of the functions coordinated by IN during the viral replicative cycle.

Natural history of HIV-1

HIV-1 is a member of an extensive and diverse family of viruses called *retroviridae*, enveloped diploid RNA viruses that are distinguished by the utilization of two unique replicative strategies: reverse transcription of the RNA genome into linear, double-stranded DNA, and the subsequent integration of this DNA into the host cell's chromosomal DNA. Prior to the discovery of HIV-1 as the etiological agent of AIDS in 1983, there was only one other retroviral infection known to cause disease in humans, the Human T-cell Leukemia Virus (HTLV-1 and II). Initially, some scientists believed that HIV-1 was a new strain of HTLV, and was nominally designated as HTLV III, but subsequently, studies of virion morphology and genetic composition indicated that this novel retrovirus (Lymphadenopathy Virus, LAV, 1983) was similar to a distinct subgroup of retroviruses called the lentiviruses [6, 7]. Lentiviruses are so named as an indicator of the characteristics of the disease they induce, a typically long incubation period between initial infection and the onset of serious symptoms, and a chronic, persistent infection (*lentus*, latin for 'slow'). Lentiviruses are now known to infect a number of non-human species, including horses (equine infectious anemia virus, EIAV), goats (caprine arthritis-encephalitis virus, CAEV), sheep (Visna virus), cats (Feline Immunodeficiency Virus, FIV), and monkeys (simian immunodeficiency virus, SIV), a common feature of the diseases they bring about in their natural hosts being the loss of function or killing of specific cells and tissues, even in the face of persistent and robust immune responses. In the case of HIV-1, the major cell population targeted is the $CD4^+$-

helper T cells, an arm of immune system involved in cell-mediated immunity. The gradual erosion of this cell population leads to the acquisition of AIDS.

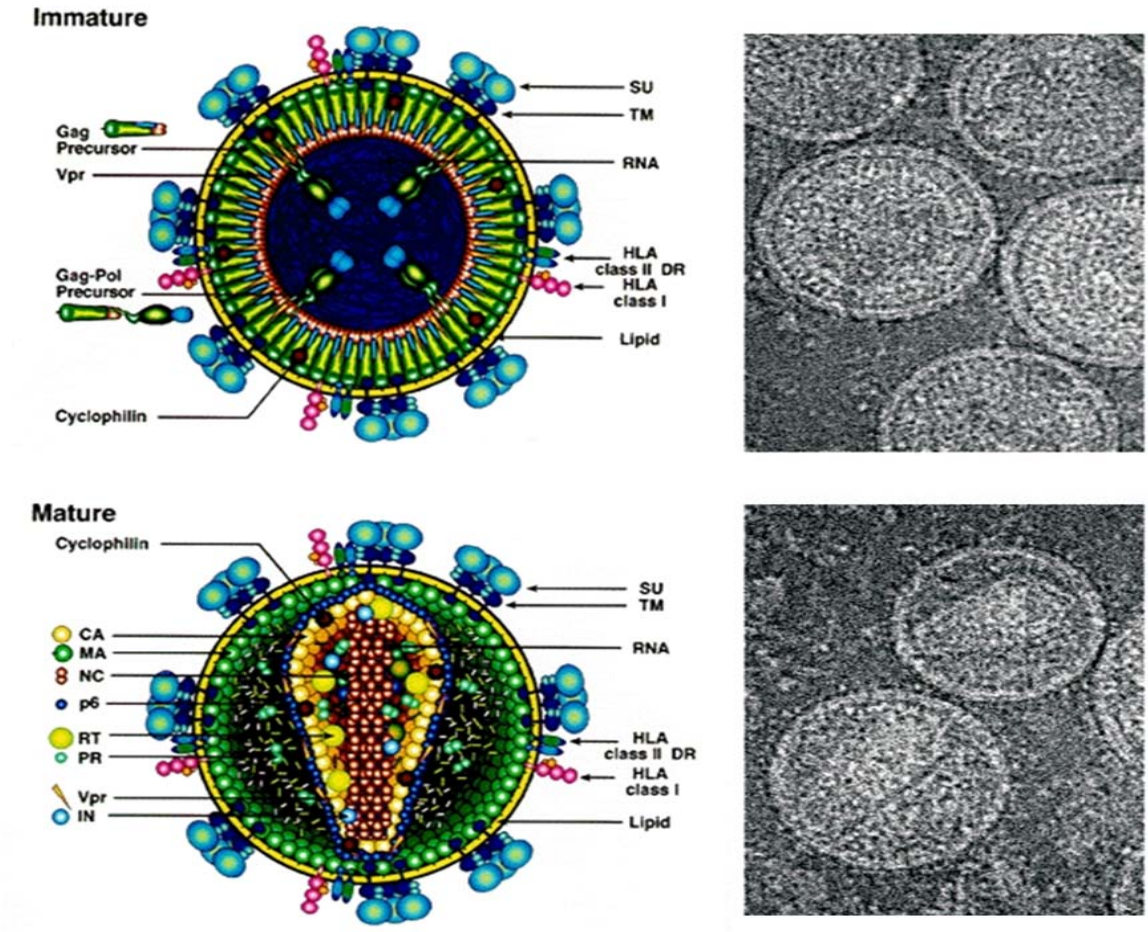
In addition to the common clinical manifestations of the disease courses, the lentiviruses share a characteristic virion morphology and genetic architecture. Under the electron microscope, the mature lentiviral particle, or virion, is observed to contain a cone-shaped core, a core shape unique to the lentivirus genus (Figure 1A). The core is comprised of approximately 1500 molecules of the viral capsid (CA), whose innate self-associative properties promote the formation of a lattice-like cage that encapsulates the diploid RNA genome and other associated viral proteins. In the virion, the core itself lies in a sequestered environment, the boundaries of which are defined by a lipid bilayer of cellular plasma membrane origin, which assumes a typically spherical shape approximately 100nm in diameter, and in which is embedded outwardly-projecting viral envelope spikes.

The lentiviral genetic architecture is an additional defining characteristic of this class of retrovirus. All retroviruses encode dedicated genes for the structural (Gag), enzymatic (Pol) and envelope (Env) protein components of their viruses. ‘Simple’ retroviruses contain no more than this minimal genetic information, whereas ‘complex’ retroviruses, which include the lentiviruses, encode several other accessory proteins that assist in various aspects of viral replication (Figure 1B). For HIV-1, these proteins include two viral factors Tat and Rev (described in greater detail in ensuing sections), which, by virtue of their absolute requirement for replication, are often precluded descriptively from the status of ‘accessory’. The virus encodes three other accessory

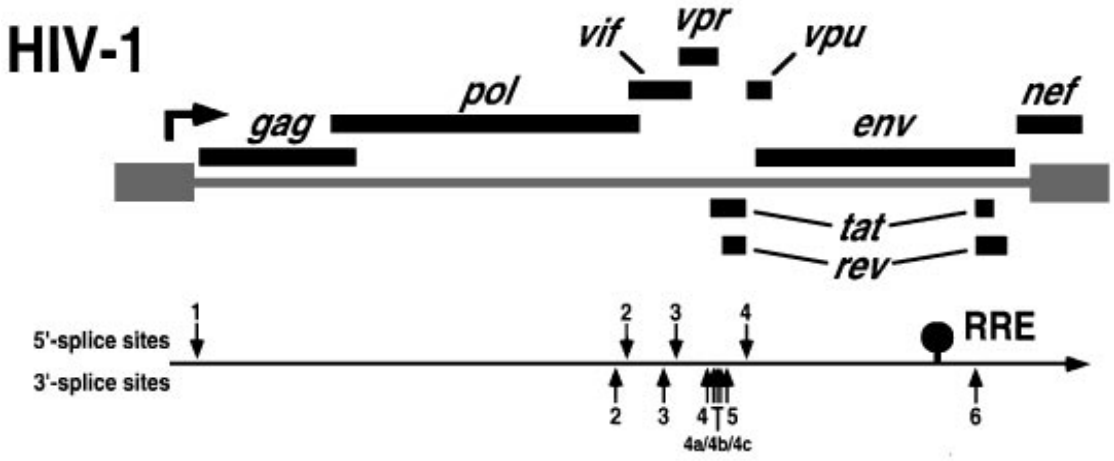
proteins: viral infectivity factor (Vif), viral protein R (Vpr), viral protein U (Vpu) and negative factor (Nef), proteins that modulate several processes to provide the ideal cellular environment for viral replication. These latter three products can more suitably be described as accessory proteins, for in some *in vitro* tissue culture models of HIV-1 infection, viral propagation can proceed efficiently, and in some cases more proficiently, in their absence. The dispensability of these accessory proteins does not however extend to *in vivo* infections, wherein the actions of some of these proteins become more essential. Overall, the accessory genes of the complex retroviruses confer a level of control over viral replication beyond that of their simple retrovirus counterparts, and are believed to in part account for some of the clinical manifestations and/or outcomes of infection.

Figure 1. Structure of HIV-1 virion and genetic organization. A) Diagrammatic cross-sectional representations of the architecture of the immature and mature forms of the HIV-1 virion are displayed alongside electron micrographs of the corresponding forms. Virions produced just after budding conform to the immature form, and require the activation of the viral PR to induce maturation, in which the viral precursor are cleaved into the constituent components and condense to produce the canonical cone-shaped core characteristic of the lentiviruses. B) Schematic of the genetic architecture of HIV-1, an example of a complex retrovirus, containing addition ORFs for the expression of various accessory proteins. The vertical displacement of the coding regions indicates the location in disparate reading frames. The splice sites that accommodate the expression of the Vif, Vpr, Tat, Rev, Vpu, Env and Nef from processing of the full-length mRNA are shown below. (Adapted from [8])

A.


Immature
SU
TM
RNA
HLA class II DR
HLA class I
Lipid
Gag Precursor
Vpr
Gag-Pol Precursor
Cyclophilin
Mature
Cyclophilin
SU
TM
RNA
CA
MA
NC
p6
RT
PR
Vpr
IN
HLA class II DR
HLA class I
Lipid

B.


HIV-1
vif
vpr
vpu
pol
gag
env
nef
tat
rev
5'-splice sites
3'-splice sites
1
2
3
4
RRE
2
3
4
5
4a/4b/4c
6

HIV-1 Replication Strategy

From the onset of infection and throughout the remaining lifetime of the host, infection by HIV-1 is characterized by persistent and robust viral replication, despite the presence of often potent host immune responses. HIV-1 is capable of infecting several cell types, including macrophages, dendritic cells, nervous system cells, and some epithelial cells, but the main reservoir for viral replication is primarily within the CD4+ helper T lymphocyte cell population. *In vivo* estimates from mathematical models of viral replication dynamics indicate that the life cycle is on average 1.2 days long [9], and results of different studies have placed the burst size, or the number of virions produced throughout the lifetime of one infected cell, from numbers ranging from $0.92x10^3$ to $5.5x10^4$ virions [10-13].

The HIV-1 life cycle (Figure 2) can be bifurcated into early and late stages. The early stage entails events that occur when a susceptible cell first encounters a viral particle, or virion, and encompasses the events of viral entry, the ensuing reverse transcription of viral RNA as the viral core traverses the cytoplasm en route to the nucleus, and culminating in the formation of the integrated viral DNA (provirus). The subsequent late stage encompasses the program of proviral gene transcription and translation, and the assembly of viral components for the eventual release of progeny virus, and their subsequent maturation. Throughout this dissertation, this designation of early and late life cycle events is used to describe the phenotypic manifestation of HIV-1 mutants, the viral mutant phenotype described as being blocked in the early or late stages of replication.

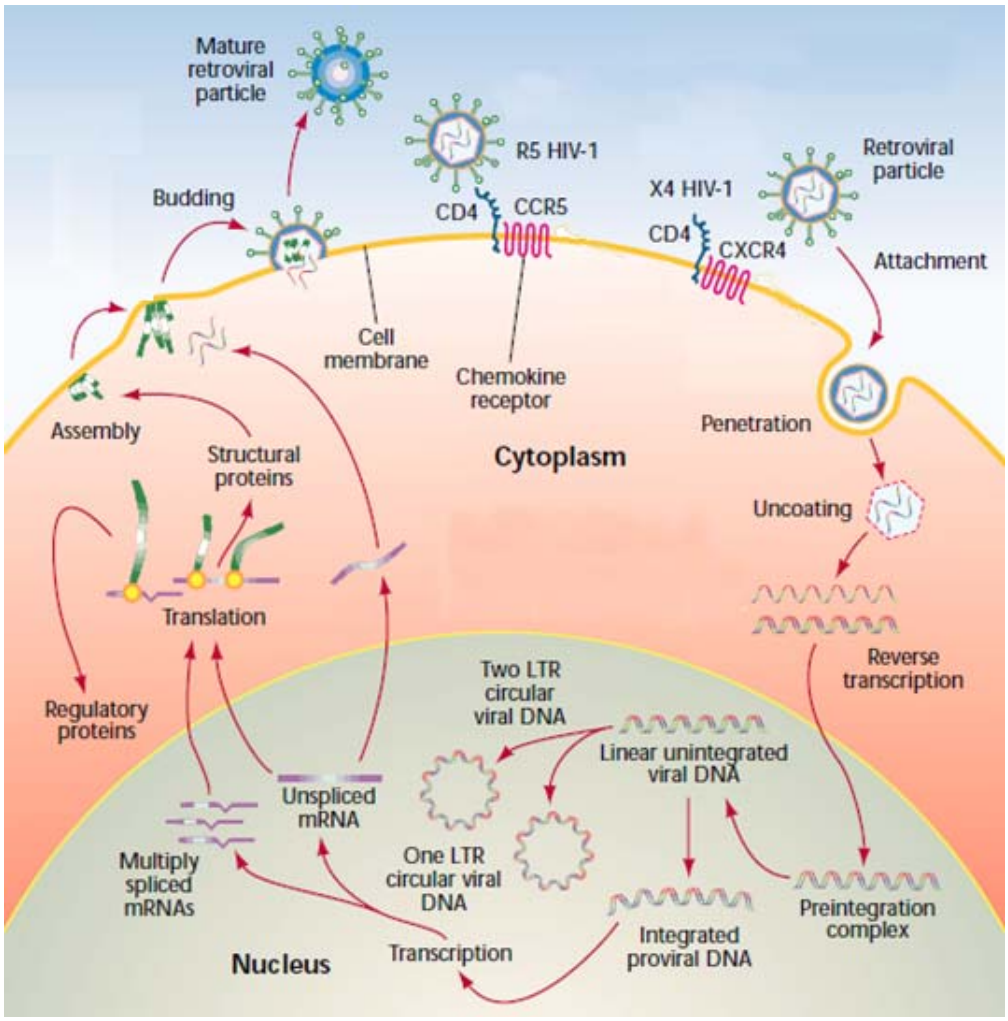


Figure 2. The HIV-1 life cycle. The early events of the viral life cycle begin with viral entry, and end with the establishment of the integrated provirus. The late events entail the program of viral gene transcription and protein translation that ultimately conclude with the production of new viral particles that begin the infection cycle anew. (Adapted from [14])

Late stage of virus replication:

Viral gene transcription and translation-

The establishment of the provirus in the host genome demarks the boundary between early and late virus life cycle events. As mentioned, the late stage encompasses the transcriptional and translational program of viral gene expression that culminates ultimately in the release of progeny virus that will begin the infection cycle anew. The organization of the HIV-1 genome as it exists in the viral RNA and proviral DNA forms are depicted in Figure 3 below.

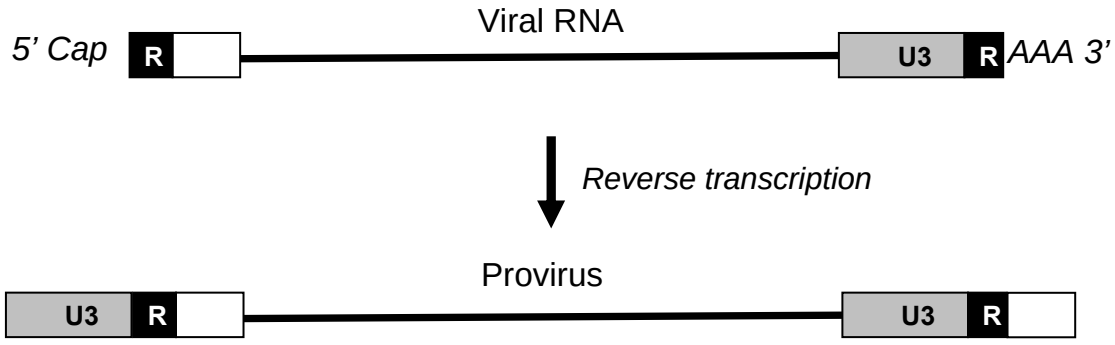


Figure 3. Viral genome in RNA and DNA forms. The process of reverse transcription creates the long terminal repeats, or LTRs, at both ends of the provirus form. This positions the transcriptional promoter and enhancer elements at the 5' end, and poly-A signals at the 3' end. R: Terminal direct repeat RNA; U5: Unique regulatory sequences at the 5' end; U3 Unique regulatory sequences at the 3' end.

The long terminal repeats, or LTRs, are direct repeat sequences that flank the proviral genome, and are formed as a consequence of the reverse transcription process

(described below). The upstream or 5' LTR contains transcriptional enhancer and promoter elements that recruit the host transcriptional machinery, and the downstream or 3' LTR contains poly-A addition signals for transcription termination and for posttranscriptional mRNA processing. The integration of the viral genome into that of the host cell makes the viral DNA indistinguishable from that of chromosomal genes, and as such the provirus is transcribed by the host's RNA polymerase II machinery.

The HIV-1 transcriptional program is an intricate process entailing a biphasic scheme of 'early' and 'late' proviral gene expression, not to be confused with the terminology used above to describe the two stages of the viral life cycle. The transcriptional scheme helps overcome two obstacles to proviral gene expression. The initial encountered barrier is the inherent inefficiency of the viral LTR to elongate the transcriptional unit [15]. In the absence of the specific regulatory viral transactivator protein (Tat), RNA polymerase II complex efficiently initiates transcription from the LTR, but then frequently pauses and disengages the DNA template soon after initiation, producing predominantly short, abortive transcripts [15]. To circumvent this but prior to Tat synthesis, basal levels of LTR activity permit the formation of low levels of full-length viral transcript, which are spliced to completion to a number of distinct iterations, determined by the utilization of different pair-wise combinations of 5' and 3' splice sites on the viral transcript. Three viral proteins, the early gene products, are translated from these fully-spliced transcripts: Tat, Rev and Nef (negative factor). Tat is targeted to the nucleus via an embedded nuclear localization sequence (NLS) [16, 17], where it binds to its cognate transactivator response element (TAR), a 59 bp stem-loop RNA structure

specified within the untranslated U5 region of the viral mRNA transcript [18-20]. Tat binding recruits several host proteins to the TAR element, the end result of which is the hyperphosphorylation of the carboxyl-terminal domain of the large subunit of RNA polymerase II, which greatly enhances its processivity on the proviral DNA template [reviewed in [21]] . In the presence of Tat, it has been estimated that the transcriptional activity of the LTR is enhanced by a factor ranging from hundreds to thousands of fold. The Tat circuitry establishes a positive feed-back loop that drives high level production of fully-spliced viral RNA transcripts, thus quickly increasing the levels of the early gene products (Tat, Rev and Nef) and completing the early phase of proviral transcription.

The second control point exists for the expression of the 'late' gene products, the structural and accessory proteins: Gag, Pol, Vif, Vpr, Vpu and Env. This nexus has its basis in the fact that only fully-spliced, intron-free mRNAs can be ushered into the cytoplasm to be translated. The viral proteins are however encoded by transcripts which contain signals (5' and 3' splice sites and somewhat ill-defined RNA structural elements) that actively retain and in some cases promote the decay of these RNAs within the nuclear compartment [22-24]. To switch on the expression of the late viral genes, the virus relies on the accumulation of the early gene product, Rev. In effect, Rev is required to promote the transport of the late mRNAs into the cytoplasm where they can be translated. However, Rev must attain a sufficiently high nuclear concentration before cooperative and functional binding by at least 8 Rev molecules to an intricate RNA structure located within the *env* gene, the Rev-responsive element (RRE) [25-27]. The RRE is found in full-length and partially spliced viral transcripts (Figure 1B), and Rev-

binding facilitates the active transport of these transcripts out of the nucleus and promotes their cytoplasmic accumulation and at later stages, full-length viral genomic RNA makes it to the plasma membrane where it is packaged into new virions. The partially spliced mRNA species contain ORFs for the accessory proteins Vif (virus infectivity factor), Vpr (viral protein R), Vpu (viral protein U) and the envelope protein (Env), whereas the full-length unspliced transcripts directs the expression of the Gag and Gag-Pol polyprotein precursors.

Assembly-

The *gag* gene encodes the 55kD precursor protein ($Pr55^{Gag}$), a conglomerate of the main structural proteins of the virion. The precursor is cleaved during or just after virus by the viral protease into its constituent parts: matrix (MA p17), capsid (CA p24), nucleocapsid (NC p7), and p6, each of which, with the exception of p6, has its individual role within the viral particle itself. Prior to budding, however, the Gag precursor polyprotein is the minimal viral component required for directing the assembly and release of virus particles. Indeed, exogenous expression of Gag, in the absence of all other viral constituents, is sufficient to assemble the production of virus-like particles (VLPs) [28-30]. A co-translational modification that appends a myristic acid moiety to the N-terminal MA domain directs precursor protein-plasma membrane association, and facilitates the intimate involvement of Gag with the membrane's inner leaflet as viral components are assembled and budding commences [31]. The Gag p6 domain, located at the carboxyl-terminus of the precursor, contains a short amino acid motif (Pro-Thr-Ala-

Pro [PTAP]) known as a late domain, which interacts with components of the machinery that comprise the multivesicular body pathway that normally recycles vesicles in the endosomal membrane system [32, 33]. HIV-1 usurps the components of this pathway for its own ends, the assembling Gag polyproteins inducing outward membrane curvature, whilst host protein constituents of the multivesicular pathway subsequently accommodate pinching off of nascent particles from the cell surface.

The *pol* gene encodes the enzymatic viral proteins [viral protease (PR p10), reverse transcriptase (RT p66/p55) and integrase (IN p32)] which are co-packaged into virions in the context of the large Gag-Pol precursor polyprotein ($Pr160^{Gag\text{-}Pol}$). The translation of both the Gag and Gag-Pol precursors are derived from the same transcript, the virus utilizing a unique method of accessing the *pol* ORF, which is encoded in the third reading frame of the mRNA transcript relative to the *gag* sequence. During translation of the viral transcript, the ribosome encounters a 'slippery' sequence, a string of consecutive uracil residues. This 'slippery' sequence, in combination with a nearby downstream RNA secondary structure, causes the ribosome to momentarily pause translation and slip back a single nucleotide on the mRNA before recommencing translation, thus shifting into the *pol* ORF, and translating the Pol products [34, 35]. The *pol* gene is thus translated as Gag/Pol fusion protein with the majority of the Gag protein appended at the N-terminus of the Gag-Pol precursor. Ribosome slipping occurs approximately 5% of the time, thus accounting for the ratio of 1 Gag-Pol precursor for every 20 Gag precursors found in the cell and subsequently in the virion. This delicate balance of Gag to Gag-Pol product is important for precision of virus assembly, for it has

been shown that overexpression of Gag-Pol can inhibit virion production [36]. The Gag-Pol precursor, as it contains Gag sequences at the N-terminus can consequently be myristylated, and intermolecular interactions between the shared Gag domains of the $Pr55^{Gag}$ and $Pr160^{Gag\text{-}Pol}$ are believed in part to direct the Gag-Pol fusion protein to assembling virions [37, 38].

The envelope protein (Env) is translated as a 160kD precursor by the protein translation machinery of the rough endoplasmic reticulum (ER), which directs the protein into the secretory pathway for its eventual delivery to the cell plasma membrane. During its egress through the membranous compartments of the ER and Golgi apparatus, the protein is extensively glycosylated, and at some point is cleaved by host proteases into the final gp41 and gp120 components of the mature envelope protein [39, 40]. gp41 is the portion of envelope that is imbedded in the plasma membrane, and is referred to as the transmembrane (TM) component of the envelope. The gp120, referred to as the surface (SU) env segment, lies on the cell exterior, and subsequently the outside of the virion, and remains associated with gp41 via non-covalent interactions. Once deposited on the plasma membrane, env diffuses laterally across the plasma membrane and interacts with assembling Gag precursors on the inner leaflet of the plasma membrane. On the virion, the viral Env assumes a trimeric formation, and determines the viral tropism, facilitating entry into select target cells that bear the required receptors that can engage the specific class of HIV-1 envelope protein encountered.

The final significant viral component, the RNA genome, is packaged at a quantity of two positive-sense strands per particle. These RNA molecules are post-

transcriptionally modified as natural host mRNA transcripts, with a 5' cap and an appended 3' poly-A tail. The packaging signal, or psi element (ψ), an extensive RNA secondary structural element specified at the 5' end of the viral transcript, mediates the association of the genomic RNA with the NC domain in the Gag precursor at the plasma membrane [41-43].

Budding and Maturation-

As the Gag and Gag-Pol precursor polyproteins assemble on the inner surface of the plasma membrane, membrane curvature is induced by Gag multimerization on the inner surface of the plasma membrane. Viral particles bud from the plasma membrane of the cell, a process that requires the activity of the viral protein U (Vpu) [44-46]. The virion, as it is released, is in an 'immature' form, and is morphologically distinguishable from the mature particle counterpart. Immature particles are observed under the electron microscope as possessing an electron-dense cloud of material just on the inner portion of the membrane, which is presumptively the closely-packed, membrane-associated precursor proteins, and an electron-lucent center (Figure 1A). Maturation of the virion coincides with the activation of the viral protease (PR), which cleaves the precursor proteins into the constituent, mature proteins. The timing of protease activity needs to be precise; premature activation, as observed in some IN mutants, releases mature viral proteins into the cytoplasm and prevents their packaging into virions [47, 48]. Furthermore, since virion assembly requires the domain activities of the extreme N- and C-termini of the intact Gag precursor, premature proteolytic cleavage also interferes with

virion assembly. The mechanism of the switch that controls PR activation is not well-understood, but it is apparent that Gag-Pol dimerization is a prerequisite [49]. The sequestered environment of the virion is thought to increase the precursor protein concentration in the immediate environment, which in turn promotes the dimerization leading to PR self-cleavage, and the inexorable processing of the surrounding precursors. The cleaved CA proteins from the precursors condense to assume the characteristic cone-shaped viral core, in which is enclosed the viral RNA and other viral cleavage products including RT and IN. The RNA itself maintains an association with NC protein, which contributes a protective effect on the viral genome. The free MA domain remains associated with the plasma membrane, and provides support for the virion superstructure [50]. The mature virion is now infectious and capable of perpetuating the viral life cycle by infecting a new cell.

Early stage of virus replication:

Entry-

HIV-1 preferentially replicates in CD4+ T cells, but will enter any cell, human or otherwise, that expresses on its cell surface the required receptors. The viral glycoprotein envelope that decorates the surface of the virion is the determinant of host range and cell tropism, as appropriate and specific engagement with uniquely expressed cell surface receptors on the target cell is a prerequisite for activating the membrane fusogenic activity of the envelope [51-54]. HIV-1 entry is dependent on interactions between the envelope protein and two distinct receptors on the cell surface. One receptor is invariably

the CD4 glycoprotein, and it is to this that the initial binding occurs with the gp120 SU subunit of Env. This results in a conformational change in gp120 that facilitates its binding to its required co-receptor, which can be either the CXCR4 and CCR5 chemokine receptor. These chemokine receptors are multipass membrane-spanning proteins involved in G-protein signaling [55, 56], and *in vivo*, are expressed on defined and distinct cell populations. The envelope protein is usually able to engage only one or the other co-receptor, the determinants for co-receptor utilization lying in the variable loop 3 (VL3) segment of the gp120 subunit [57]. Since CXCR4 is co-expressed with CD4+ mostly in T-helper cells, viruses that utilize the CXCR4 receptor are called T-tropic viruses. CCR5 is expressed with CD4 on both CD4+ T cells and macrophage. Viruses using CCR5 as a co-receptor for entry were originally classified as M-tropic viruses.

After engagement of the envelope with the receptor and co-receptor, structural rearrangements of the gp120 subunit activate the gp41 TM subunit, which contains a fusion peptide. In the context of the unengaged envelope the fusion peptide sequence is buried within the trimeric structure. However, upon receptor binding, internal molecular rearrangement of gp41 SU results in the insertion of the fusion peptide into the cellular membrane, an action that juxtaposes the viral and cellular membranes. The two membranes fuse, uniting the intravirion and intracellular compartments, and the viral core is ushered into the host cell's cytoplasm.

Reverse Transcription-

After entry, certain components of the macromolecular complex of the viral core disassemble in a little-understood process called uncoating, leaving the viral RNA in complex with a subset of the original core proteins in the cytoplasm, or just outside the nuclear envelope. This ribonucleoprotein complex is called the reverse transcription complex (RTC) or preintegration complex (PIC), and it is in this context that the program of reverse transcription occurs [58]. The PIC is believed to contain the viral RT, IN, NC, Vpr, and other associated host proteins, but an accurate depiction of its components is lacking.

RT is an RNA- and DNA-dependent polymerase, and functions as a heterodimer of two related subunits, the p66 and p51 subunits. The p66 subunit contains the RT domain, and an RNaseH domain appended at the C-terminus. The p51 subunit is derived from the p66 subunit, but with the RNaseH domain removed by HIV-1 protease action. This latter subunit does not contribute any catalytic activity, but instead appears to contribute a structural role for the assembled heterodimer [59]. A tRNA molecule that is selectively packaged in the virion, $tRNA_3^{Lys}$ [60], hybridizes with a complementary sequence on the genomic RNA close to the 5' end called the primer binding site (PBS), and primes the initial synthesis of minus-strand DNA, which extends to the 5' end of the molecule [61]. This initial synthesized DNA is called minus-strand strong stop DNA, and contains a sequence found at both the 5' and 3' ends of the viral RNA called repeat (R) sequences that are direct repeats of each other. The RNaseH domain of RT digests the RNA in the RNA-DNA hybrid, and frees the minus-strand strong stop DNA to hybridize

with the R sequence at the 3' end of the genome. This template switching is called strand transfer (not to be confused with IN strand transfer), and is the first of two such events to occur during reverse transcription.

After the strand transfer, minus-strand DNA synthesis proceeds from the 3'end, with the RNaseH domain of RT again digesting the RNA in the lengthening hybrid as elongation occurs. Certain purine rich regions of the genomic RNA, one located at the 3' end of the genome (3' polypurine tract or 3' PPT) and another close to the center of the genome within the IN region (central polypurine tract or cPPT), are less-susceptible to RNaseH digestion however, and as the minus-strand synthesis proceeds, the RNA in these regions persist and serve as primers for the synthesis of plus-strand DNA in the opposite direction. The first plus-strand DNA synthesized terminates at the PBS sequence on the minus-strand strong-stop DNA from the first strand transfer jump, and is called plus-strand strong-stop DNA. The plus-strand strong-stop DNA then facilitates the second strand transfer event, hybridizing with the PBS at the 5' end. Elongation then proceeds in both directions from both DNA strands, resulting in the formation of the long terminal repeats (LTRs), and the double-stranded cDNA, which is the substrate for the viral IN. Reverse transcription is an error-prone process, since RT possesses no proof-reading activity, and accounts partly for the high mutagenic rate of the virus. It is estimated that the mutation rate is on the order of 3×10^{-5} to 4×10^{-5} mutations per target base pair per cycle [62, 63].

Nuclear translocation and integration:

The lentiviruses, such as HIV-1, are unique in the regard that they are fully capable of infecting terminally differentiated, non-dividing cells, or cells that are arrested in the cell cycle [64-66]. Since integration of the viral genome by necessity requires the juxtaposition of viral components and the chromosomal DNA, the permanently intact nuclear membrane in these cellular targets presents a natural barrier to the cellular DNA. Indeed, some retroviral infections, for example those of Moloney murine leukemia virus (Mo-MuLV) and Rous sarcoma virus (RSV) are dependent on the progression of the cell cycle into mitosis, wherein the dissolution of the nuclear membrane provides access to the cellular DNA [67-69]). HIV-1 in contrast is known to productively infect terminally differentiated macrophage and dendritic cells. The HIV-1 PIC engages components of the nuclear pore complex (NPC), a protein-lined channel that spans the nuclear double membrane, which regulates the bi-directional molecular traffic between the cytoplasm and nucleoplasm, and can be transported in an energy-dependent manner into the nucleus [70] . It is currently a matter of intense debate as to the specific determinants in the PIC that facilitate its transport into the nucleus. Nuclear localization signals (NLS) have been found in Vpr [71], IN [72], and MA [73] proteins, and there are even some implications that a central DNA flap [74], which is a portion of overlapping DNA sequence created at the cPPT during reverse transcription, could facilitate nuclear entry. Whichever may be the case, it is clear however that the macromolecular deoxyribonucleoprotein complex constituting the PIC exceeds the upper limit (50kDa) for passive diffusion through the

nuclear pore [58], and thus depends on active facilitated transport mechanisms to get it or parts thereof into the nucleus.

The integration reaction will be described in more detail in further sections, but a brief overview will be given here. The integration reaction itself is divided into two sequential and spatially separated reactions, the first of which putatively occurs in the cytoplasm upon completion of reverse transcription. This reaction, called 3' processing, proceeds with IN recognition and binding of the viral cDNA ends, and the removal of a terminal dinucleotide from the 3' end of both DNA strands. This activates the DNA ends for the subsequent strand transfer step, which occurs once the PIC has traversed the nuclear pore. Here, IN covalently links the recessed 3' DNA ends to the chromosomal DNA in a concerted cleavage-ligation reaction. The immediate product is the viral DNA covalently linked at both 3'ends, but with a 5 bp stretch of unpaired bases flanking the site of insertion. Host DNA repair enzymes are then believed to fill in the unpaired gaps at either end of the insertion. The viral genome thus becomes contiguous with and undistinguishable from cellular genes, and can be transcribed by the cellular RNA polymerase II machinery. Though integration can potentially occur anywhere in the genome, HIV-1 exhibits a predeliction to integrate in chromosomal regions where genes are actively transcribed, where the decondensed chromatin structure of the immediate environment is more conducive to efficient gene transcription.

Alternative fates of viral DNA:

Integration of viral cDNA into chromosomal DNA is an essential step in the retrovirus life cycle, and is considered to be the natural fate of the newly reverse-transcribed DNA product. Integration is, however, only one of several fates for viral DNA. Even in wild-type (WT) viral infection, significant levels of unintegrated linear and circular viral DNA forms can be detected [75-77]. The circular forms contain either one or two copies of the viral LTRs, and are called 1-LTR and 2-LTR circles respectively (Figure 4). The mechanism by which 1-LTR circles are produced is uncertain, but they are believed to be either the product of errors incurred during the reverse transcription reaction [78, 79], or created by homologous recombination between the two LTRs after nuclear entry [80, 81]. 2-LTR circles are purportedly produced by the action of resident nuclear DNA ligases that join the free blunt-ended viral DNA ends together [82]. Since ligation activity is localized to the nuclear compartment, the formation of 2-LTR circles is often used as a surrogate marker for the successful nuclear import of viral DNA. Quantification of 2-LTR circles is also used as an in vivo measure of the catalytic efficiency of IN mutants. IN mutants deficient for integration catalytic activity, which includes the initial processing of the viral DNA ends, produce a preponderance of blunt-ended viral DNA species, an ideal substrate for blunt-ended DNA ligation to produce high levels of 2-LTR species. The 1- and 2-LTR DNAs are considered dead-end viral DNA products of reverse transcription, for the transcriptional activity from these species is very limited, and, possessed of no other means of replicating, they are eventually diluted out of replicating cell cultures after several rounds of cell division. In non-

replicating cells they are quite stable, however, as circular DNA forms can be detected in infected macrophage cultures up to 30 days post infection [83]. The role, if any, of these circular forms in HIV-1 pathogenesis is still uncertain.

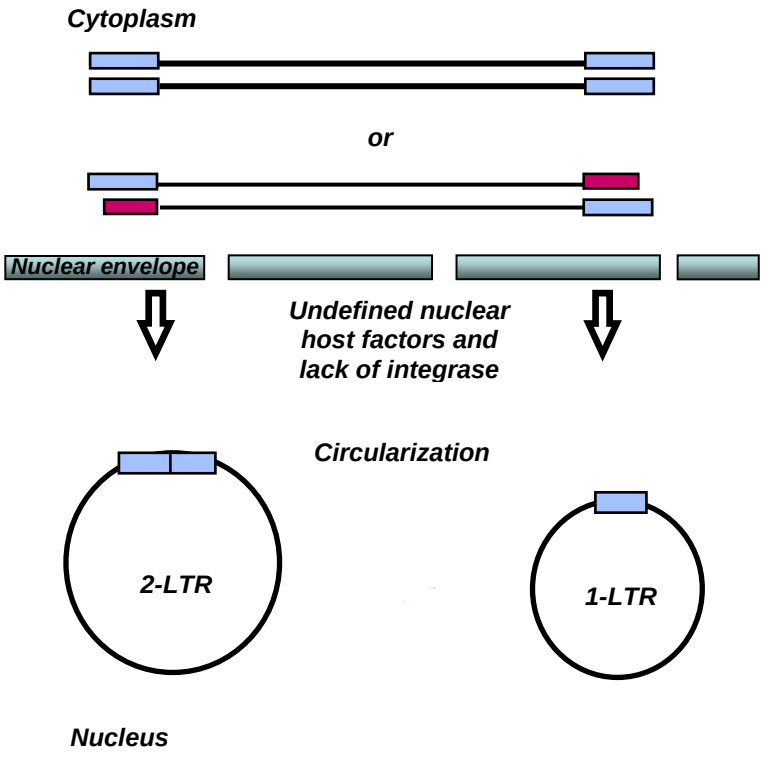


Figure 4. Alternative fate of viral DNA. Reverse-transcribed viral cDNA is the required precursor of the integrated provirus, but in the absence of IN catalysis, circular products containing one or two copies of the viral LTR sequences (1- and 2-LTR circles respectively) accumulate in the nucleus. Since 2 LTR circles are contingent upon nuclear ligases for their formation, their presence is indicative of successful nuclear import of viral cDNA.

Roles of Accessory Viral Genes:

The preceding overview of the HIV-1 life cycle describes only the minimal processes required for the viral replication program. In many *in vitro* infection models using immortalized cell lines, these processes are sufficient for viral replication, and in such circumstances the accessory proteins Vif, Vpr, Vpu and Nef are dispensable. *In vivo* replication, however, presents several challenges to the virus not encountered *in vitro*, and it is here that the actions of most of the accessory proteins are required to surmount these host-mediated barriers to replication. A comprehensive review of these accessory proteins will not be provided here, but their basic functions will be outlined [for general review see [84]]. This will provide a framework for understanding the selection of certain genetic backgrounds used for specific assays, in which the activity of an accessory gene can be detrimental for the purposes of the particular assay measurement.

Nef-

Nef is one of the early viral gene products, expressed in abundance soon after establishment of the proviral DNA. The protein is a small (25 kDa) membrane-associated protein, utilizing an attached myristylic acid to anchor its N-terminus into the plasma membrane. Though Nef has been attributed with a multitude of functions, including enhancement of virion infectivity [85], inhibition of the apoptosis [86], and up-regulation of specific surface markers on distinct cells [87], the most recognized ones are the downregulation of CD4 and MHC-class I molecule on the cell surface [88, 89]. The removal of CD4 molecules from the cell surface increases the efficiency of viral release

as CD4 can interact with the viral envelope on coalescing virions on the cell surface [90]. MHC-I downregulation functions as a means of evading immune system surveillance by cytotoxic T lymphocytes (CTL), which scan the surfaces of cells for the presence of foreign peptides which are loaded onto MHC-1 molecules. These activities are obviously not required in the tissue culture dish, and the dispensability of Nef in combination with the propitious location of its ORF at the end of the genome, where there is little overlap with other protein ORFs, make this region of the very compact HIV-1 genome amenable to genetic manipulation. The insertion of genetic markers such as GFP or drug resistance markers into the readily replaceable *nef* gene is commonly used as an investigative tool for tracking viral infectivity and gene expression.

Vif-

The major function of Vif is to subvert an innate cellular antiviral activity which is mediated by APOBEC3G/F. APOBEC3G/F are cytidine deaminases, found usually in the cytoplasm [For review see [91, 92]] . In cells that are productively infected with HIV-1, APOBEC3s are packaged with the outgoing virion, remaining associated with the core components as a new cell is infected. The antiviral activity of APOBEC3G/F is manifested during reverse transcription, wherein they potently deaminate cytidine residues on the minus-stranded DNA, leading to the significant accrual of deleterious mutations in the viral genome. Vif counters this antiviral activity by targeting APOBEC3G/F for proteasomal degradation in the producer cell, before it can be packaged into outgoing virions. Cells that naturally express low levels of APOBEC3G/F are called ‘permissive cells’, as viral replication can proceed undeterred even in the

absence of Vif. Non-permissive cells produce APOBEG3G/F in sufficient quantities that necessitate the action of Vif.

Vpr-

Vpr is a basic 14 kDa protein produced during the late transcriptional phase of viral gene production. It is found ubiquitously throughout the infected cell, but tends to accumulate in the nucleus, or associate with the nuclear membrane via nuclear pore interactions [93] [94]. Vpr is also packaged into virions, via interactions it establishes with the p55 Gag precursor [95]. For its diminutive size, Vpr has been found to elicit an array of distinct and often contradictory activities. There is however little consensus as to the actual role or roles during infection. Vpr has been demonstrated to be involved in the accuracy of reverse transcription [62, 63, 96], translocation of the viral DNA in the context of the PIC [71], regulation of apoptosis [97, 98], transactivation of the viral LTR [99], controlling transcription from episomal DNA [100, 101], and control of cell cycle progression [102].

Despite this dizzying array of proposed activities, its function is dispensable for productive infection in many immortalized cell lines, and for the infection of primary CD4+ T lymphocytes. It does however appear to be necessary for productive infection of macrophages [103]. The role of Vpr in control of cell cycle progression and a possible pro-apoptotic activity is especially pertinent in some of the assays presented here. An assay for integration frequency, which depends on selection and multiplication of cells that are successfully transduced with a dominant selectable drug resistance marker, does

not work if the virus expresses an active Vpr. It has been demonstrated that Vpr causes a G_2 cell-cycle arrest [102], and though there is evidence that the viral LTR is most active during this phase of the cell cycle [104], the prevention of cell division is detrimental for the purposes of the assay. The basis for this dilemma being the integration frequency in this assay is determined by macroscopic observation and scoring of discrete resistant colonies, each of which represents a single stably transduced cell that must undergo multiple replication cycles before macroscopic perception is manageable. It cannot however be ruled out that the purported pro-apoptotic activity of vpr might also influence the utility of this assay.

Vpu-

Vpu is a small protein that is embedded into the plasma membrane via a single-pass transmembrane helix. It was long known that this protein enhances viral release from the plasma membrane, but it was only recently discovered how it carries out this function. Vpu antagonized the action of a plasma membrane protein aptly named tetherin, which holds onto virus particles that have already pinched off from the membrane and prevents their release from the cell surface [45]. Though the mechanism by which it inhibits tetherin activity is not known, it is clear that in cells that express tetherin, Vpu is required for proper viral release. It is noteworthy that CEM cells, a non-adherent human T-cell line, is apparently preferentially infected by cell-cell transmission, as opposed to being infected by free virions. We have observed, as others have before, that a *vpu*- virus is able to significantly enhance the inherent infectivity of virus in these cultures, and that

multinucleated membrane bound entities called syncytia, produced by cell-cell fusion, is also enhanced by the absence of the Vpu protein [105].

HIV-1 Integrase

Expressed initially as the C-terminal domain of the *pol* gene (Figure 5), IN is packaged into budding viral particles in the context of the large Gag-Pol precursor polyprotein. The mature 32kDa, 288-amino acid IN is produced, as is the case with all mature *gag* and *pol* gene products, within the confines of the virion by the action of the viral protease. After viral entry, reverse transcription of the viral RNA produces the substrate for integration: a linear, double-stranded DNA copy of the viral genome. The catalytic activities of IN then orchestrate the unique replicative strategy of the retrovirus family, the insertion of the viral cDNA copy into the genome of the host cell (Figure 6). The ultimate result of IN catalytic activity is the covalent and irreversible insertion of the viral genome into the host chromosomal DNA, making it indistinguishable from the genetic information of the cell. Since efficient viral transcription only occurs from the proviral DNA form [106], integration is a crucial process for the propagation of the retrovirus, and is the underlying basis for the perpetuation of the viral genetic information throughout the lifetime of the infected cell, and to successive generations of daughter cells after mitotic cell division.

The investigation of HIV-1 IN during infection has presented an especially difficult challenge due in part to the multi-functionality of the protein. Apart from its unique role in integration, IN is required for other distinct viral processes during many stages of the viral life cycle [107], impacting the precision of proteolytic processing of the Gag and Gag-Pol polyproteins, and particle morphogenesis in producer cells [47] as well as the efficiency of reverse transcription in recipient target cells [108-113]. For

clarity, a standardized nomenclature has been derived for describing IN mutant phenotypes as either Class I or II [114]. Class I IN mutant viruses are efficiently produced and otherwise infectious proceeding through all stages including reverse-

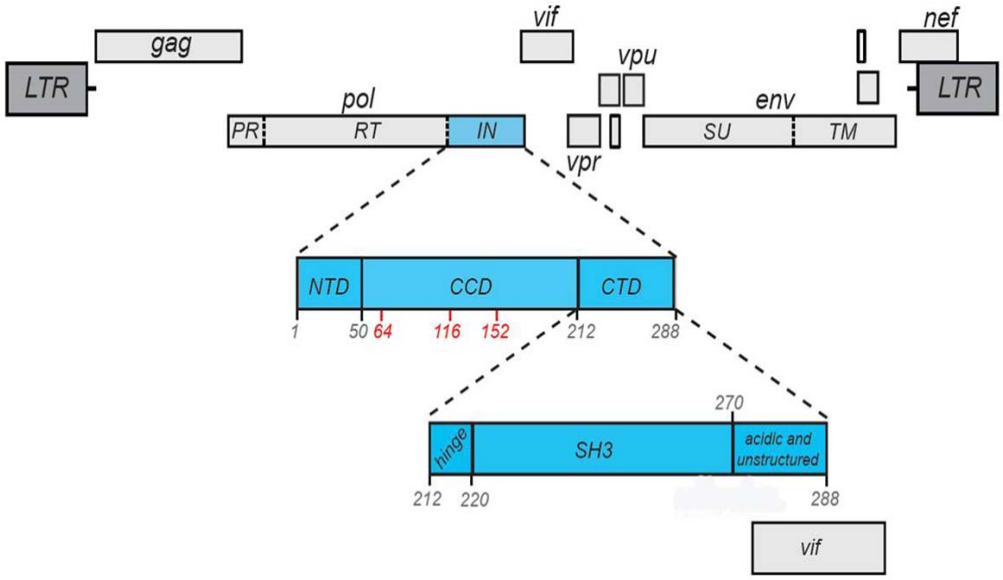


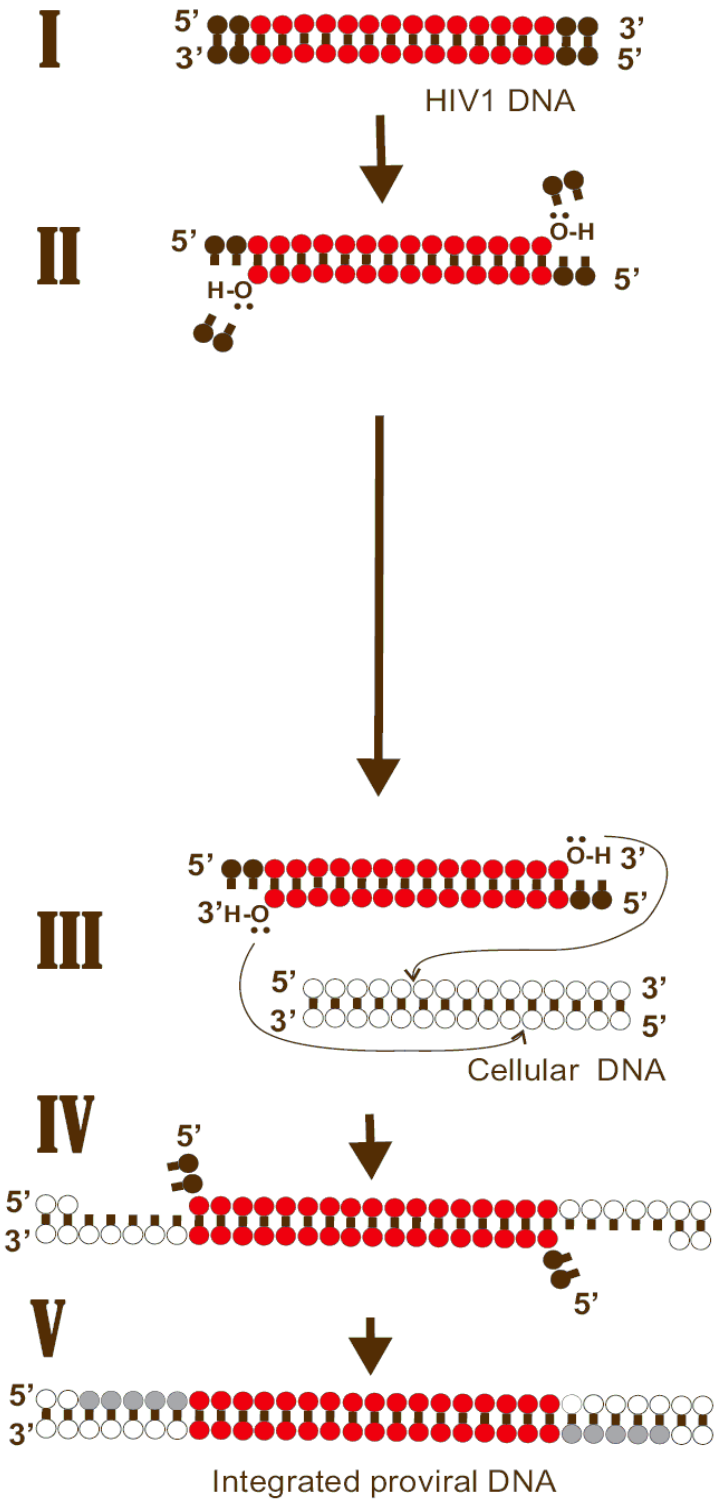
Figure 5. Position of IN in the HIV-1 genome. The nine ORFs of HIV-1 are depicted as grey boxes. The vertical displacement of each box indicates the relative reading frame. The IN region is blown up to show the domain structure, and the CTD is further blown up to display the finer structural detail. The region of overlap is also shown. The sizes of the reading frames are not drawn to scale.

transcription and viral DNA nuclear import but are blocked for the catalytic functions of 3' processing and strand-transfer. Thus, Class I mutant infections result in unprocessed, blunt-ended linear viral DNA intermediates that when transported into the nuclear compartment are expeditiously self-ligated to produce circularized viral DNA molecules (2-LTR circles), the accumulation of which is a signature feature of this mutant

phenotypic class. In contrast, Class II IN mutants, a broader phenotypic designation, manifests a range of pleiotropic effects including defective polyprotein processing, virion maturation and egress in producer cells with the potential for concomitant effects during the infection of target recipient cells, the most obvious of which is the perturbation of reverse transcriptase activity.

Figure 6. Overview of HIV-1 integration reaction. I) HIV-1 IN encounters its substrate produced from reverse transcription, a linear double-stranded copy of the viral genome. II) IN removes a dinucleotide from the 3' ends of both strands, in the first step of integration called 3' processing. III) The processed DNA, in the context of the PIC, encounters target DNA. IN directs the exposed –OH groups from the processed DNA ends to nucleophilic attack of phosphodiester bonds in target DNA separated by 5bp on opposing strands. IV) The immediate product, viral DNA linked covalently at the 3' ends to target DNA, with flanking sequences of unpaired DNA 5bp in length, and a 2bp 5' unpaired overlap of viral DNA. V) Unpaired gaps are repaired by host DNA repair enzymes. The proviral DNA is flanked by a duplicated 5bp sequence, characteristic of HIV-1 integrase-mediated integration. (adapted from [115])

INTEGRASE


I
5'
3'
3'
5'
HIV1 DNA
II
5'
Ö-H
H-O
5'
III
5'
Ö-H 3'
3'H-O
5'
5'
3'
3'
5'
Cellular DNA
IV
5'
5'
3'
5'
V
5'
3'
Integrated proviral DNA

Integrase Structural Domains:

The replication strategies of retroviruses and transposable elements entail essentially parallel processes, both requiring the mandatory step of integrating the element's genome into a foreign DNA molecule. Indeed HIV-1 IN and the transposase enzymes of DNA transposons and retroelements belong to a superfamily of polynucleotidyltransferases, which includes Mu transposase, and the *Escherichia coli* (*E. coli*) enzymes RuvC endonuclease and RNaseH [116-119]. Due to its insolubility and propensity to aggregate at high concentrations, efforts to determine a high resolution structure of the isolated HIV-1 IN protein, or complexed with its substrate DNA, have thus far been unsuccessful. Nonetheless, the overall structural architecture of the enzyme has been inferred from evidence derived from distinct experiments, including analyses of the peptide products from limited proteolysis of susceptible linker sequences between domains [120], *in vitro* complementation assays [121, 122], phylogenetic comparisons of related enzymes [120], and the successful structural determination of the individual protein domains [116, 122-129]. IN comprises three domains: the N-terminal domain (NTD), the catalytic core domain (CCD) and the carboxyl-terminal domain (CTD). The structures of each of the three domains have been determined in isolation, and two-domain structures compromising the NTD/CCD and CCD/CTD have been also been solved.

N-Terminal Domain-

The NTD encompasses residues 1 through 49, and contains an invariant HHCC motif conserved throughout retroviruses and retrotransposons. This motif is involved in zinc metal co-ordination, establishing an *in vitro*-determined stoichiometry of 1 IN molecule:1 Zn^{2+} ion [130, 131]. Metal binding has been demonstrated to facilitate the correct tertiary folding of the domain, which assumes a helix-turn-helix topology reminiscent of zinc-finger DNA-binding domains [131]. The resemblance ends here however, for there is no evidence that the IN NTD can bind DNA, either in a specific or non-specific fashion. Instead, it is believed that the NTD contributes to the ability of IN to multimerize [131], which is an ability required by IN to conduct the integration reaction. One report has also implicated the role of the NTD in specific interactions with the viral DNA substrate, the lysine at position 14 identified as the specific residue [132].

Catalytic Core Domain-

The CCD maintains the highest degree of sequence homology between distinct retroviruses, compared to the other two domains of IN. This domain encompasses residues 50-212, and contains a constellation of acidic residues that create a D,D-35-E motif that is conserved throughout the integrase enzymes of all retroviruses, and with those of retrotransposons and some bacterial transposons [133, 134]. These residues co-ordinate two divalent cations required for catalysis. For HIV-1 IN either Mg^{2+} or Mn^{2+} can be accommodated for *in vitro* assays of integration function, though it is Mg^{2+} that

likely is the biologically relevant species [135]. Alteration of any of the three catalytic residues abrogates all integration activity [120, 136, 137], and infections with mutant

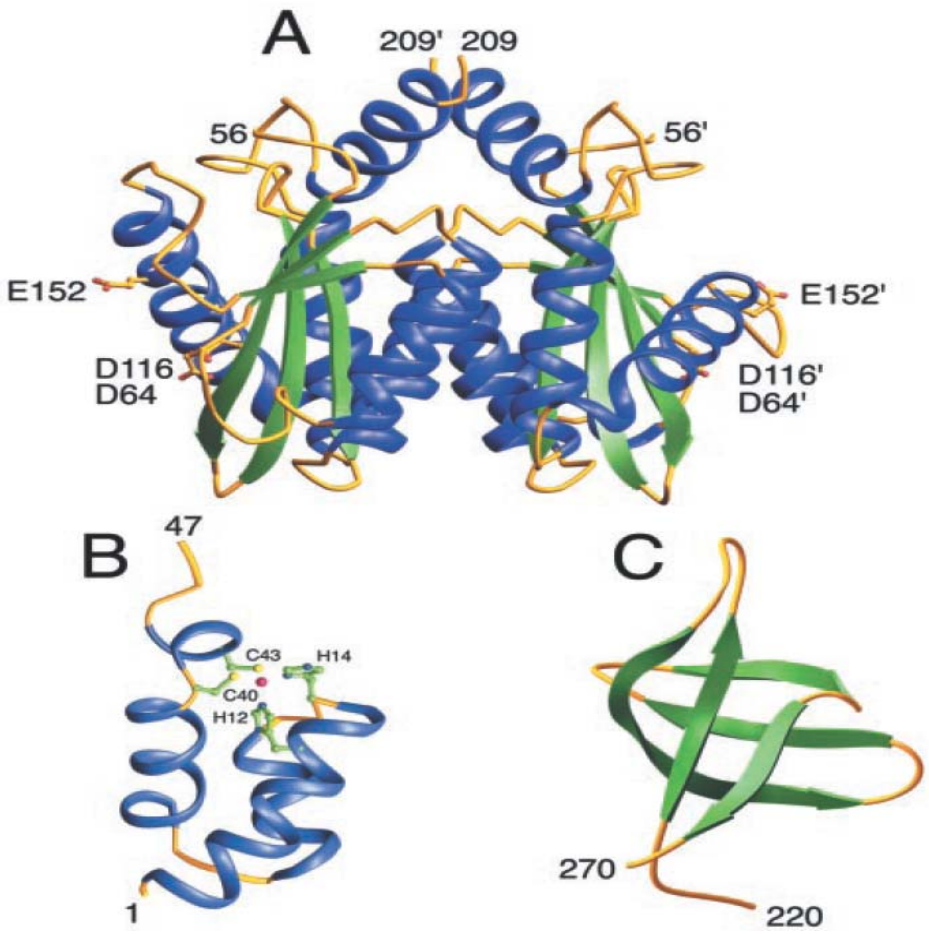


Figure 7. The domain structures of HIV-1 IN. (A) The catalytic core (CCD) structure, shown in dimeric form. The residues comprising the D,D-35-E motif are displayed. (B) The N-terminal domain (NTD) depicted with the highly conserved HHCC motif, in coordination with Zn^{2+}. (C) The carboxyl-terminal domain (CTD). The unstructured tail comprising residues 271-288 are unresolved, and thus not shown. For all structures, alpha helices are in blue, beta sheets in green. (Adapted from [138])

viruses bearing mutations at these residues characteristically accumulate high levels of unintegrated episomal viral DNA in the infected cell, and are archetypal class I mutants. The solved structure of the isolated CCD was solved as a symmetrical dimer with an extensive and likely biologically-relevant dimer interface [127]. These dimeric contacts were maintained in the crystal structure of the two domain CCD-CTD crystal, thus further highlighting its relevance [124]. The overall topology of the CCD structure assumes an RNaseH fold, and closely resembles the RNaseH domain found in the p66 subunit of reverse transcriptase. This characteristic folding pattern is maintained throughout the family members of the polynucleotidyltransferases, including the *E. coli* enzymes, RuvC resolvase and RNaseH [116]. Notably, the structure contains two flexible loops, comprising residues 138-149 and 186-194. The first loop is disordered and therefore absent from the crystal structure, and sits on the lip of a cleft which houses the catalytic triad. The presence and importance of this flexible loop to the integration reaction lends some insight into the dynamic structural flux required of the enzyme during the integration reaction.

Carboxyl Terminal Domain-

The CTD encompasses residues 212-288, and is connected to the CCD via an extended α-helical coil (residues 195 to 220). Even amongst retroviruses, the CTD exhibits minimal homology, and contains a single invariant residue, W235, common only amongst retroviral integrases. Though relatively well-conserved, the identity of the residue at the 235 position in HIV-1 IN is not critical, since the virus can accommodate

certain amino acids substitutions at this position [107]. The structure of the domain was solved by nuclear magnetic resonance (NMR) as a dimer; with the isolated monomeric unit assuming the topology of a five-stranded β-barrel that bears striking resemblance to an Src-homology 3-like fold (SH3) [125, 126]. Residues 220-270, which coincidentally are the outer margins of a non-specific DNA binding activity that is inherent to the domain, describe the boundaries of this structural element [139]. Since HIV-1 integration sites display little if any consensus sequence, it was once thought that the non-specific DNA binding activity of the CTD served as an anchor that would hold the IN-DNA complex to a random site on the chromosome. This theory has however lost merit with the discovery of IN interaction with a cellular protein, the lens-epithelium-derived growth factor (LEDGF), which is believed to act as a molecular tether between IN and chromosomal DNA [140, 141]. Instead, it is currently believed that the CTD non-specific DNA binding serves to stabilize the viral DNA substrate in such a way that it forms the appropriate contact with residues in the CCD for catalysis. Other functions ascribed to the CTD include the roles in promoting IN multimerization [142], for coordinating the binding of both host- and virion-derived proteins [3, 110, 143-148], and a possible role in PIC nuclear import [72]. It was recently demonstrated that IN and RT interact, bridged via IN CTD contacts [110, 149], and that IN enhances the efficiency of reverse transcription [109]. The CTD is especially implicated in the facilitation of secondary IN functions. Highlighting this, a mutagenic substitution study of the most conserved residues shared between HIV-1/SIV_{cpz} isolates, showed that alteration of conserved CTD residues gave rise to predominately Class II IN mutants [2].

The structural data for the CTD ends at the aspartic acid residue at 270, with the remaining 18 residues (amino acids 271-288) proving recalcitrant to conventional structural investigation due its high level of disorder. This IN 'tail' is predominantly acidic, its role(s) in IN functions currently unknown. It has recently been shown that HIV-1 IN CTD is a potent substrate for p300-mediated histone acetyltransferase (HAT) activity [3, 4] at three lysine residues (K264, K266, K273), a phenomenon subsequently demonstrated to be non-essential for sustained viral replication through an immortalized T cell line [4]. The impact of IN acetylation has however not been evaluated in primary cells, where it might play a more pivotal role. This inquiry is hindered by the overlapping ORFs of IN and Vif. Alteration of any residue after IN 271 simultaneously alters the amino terminus of Vif, a domain recently noted for its role in defining interactions with APOBEC family members [150-152]. Unlike some immortalized cell-lines, Vif function is absolutely required in the primary cell targets of HIV-1, thus the interpretation of any mutant phenotype derived from altering residues in the IN tail is complicated by the potential influence on Vif function. It is intriguing that K273 is the most robustly acetylated residue [4], and is also a residue implicated for ubiquitination that impacts IN stability [153]; it is thus frustrating to the experimenter that K273 at the very beginning of the IN-Vif overlap.

To investigate the contribution of the IN tail, and by consequence the post-transcriptional modifications therein, we created a series of truncation mutants that sequentially removed the tail of IN, and concomitantly the residues targeted for posttranslational modification. These experiments were conducted in immortalized cells

that are Vif-permissive (do not require Vif), thus removing the complicating factor of *vif* alteration. The viruses were tested in a number of assays to assess the effects of truncation on a variety of measurable viral processes. It was found that the IN tail is dispensable for viral replication in our *in vitro* systems, but we discovered a number of truncations within a very narrow window of residues that displayed strikingly dissimilar phenotypes, manifesting gains and losses of specific viral processes.

Functions of IN:

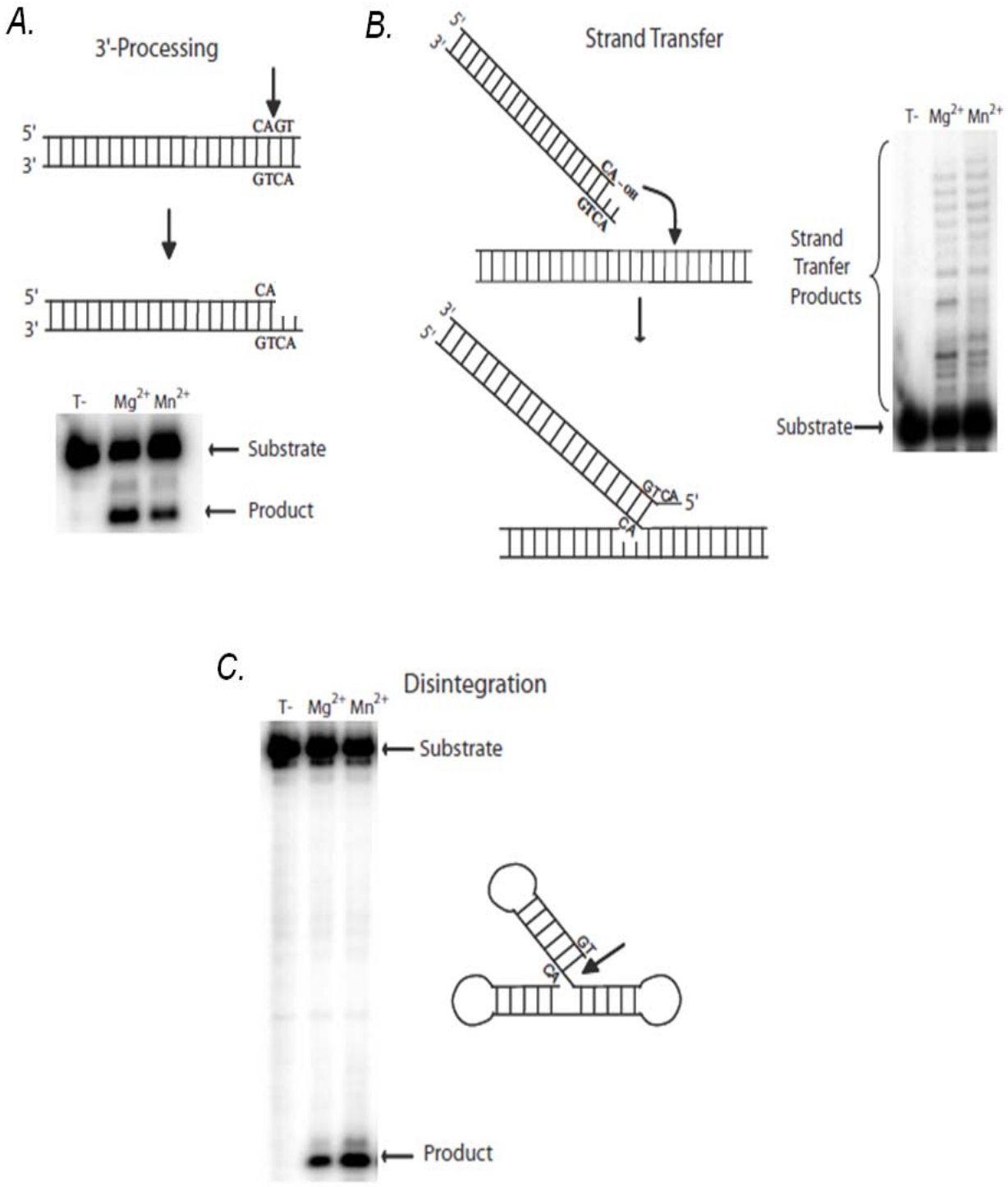
The Integration Reaction-

The primary function of HIV-1 IN is to catalyze the integration reaction. During infection, this reaction comprises two sequential and spatially separated biochemical activities called 3' processing and strand transfer. Much of the mechanistic details of the integration reaction have been deciphered from *in vitro* assays that use purified recombinant IN with DNA substrates that mimic the viral DNA ends. The 3' processing assay directly assesses the ability of IN to remove a terminal dinucleotide from a short DNA substrate with sequences from the viral DNA terminus, the resulting processed products easily resolved and distinguished from unprocessed substrate (Figure 8A).

The strand transfer reaction is a relatively inefficient process when conducted *in vitro* with recombinant IN and viral DNA substrates, highlighting the possible requirement for extraneous viral or cellular factors for this integration activity. Most reaction conditions favor the formation of 'half site' integration products (Figure 8B) derived from the insertion of only one viral end into one strand of the target DNA. 'Full

site' or concerted integration entails the successful insertion of both viral DNA ends into the target sequence, in the characteristic 5bp staggered fashion of the IN of HIV-1. These more natural products of IN integration activity can be produced *in vitro* with recombinant IN and substrate DNA alone, but the process is highly sensitive to reaction conditions, and to the concentration of IN both during purification and in the reaction itself [154]. The completion of the integration process requires the subsequent filling of the single-stranded gaps at the site of integration, a process undertaken ostensibly by cellular factors in the natural infection setting.

Figure 8. The *in vitro* activities of IN. A) 3' processing. Viral DNA substrate containing cis-acting sequences at the extreme termini, including a highly-conserved CA dinucleotide, are endonucleolytically cleaved by IN at the scissile phosphodiester bond between the CA and GT dinucleotides, indicated by the arrow. B) Strand transfer. IN directs a nucleophilic attack of a phospodiester bond on a target DNA substrate, using the exposed –OH group on the recessed end of processed DNA substrate. Most reaction conditions favor the half-site reaction, shown here. C) Disintegration. An IN-mediated processed observed only *in vitro*, it is considered a reversal of the strand transfer reaction, resolving the intermediate DNA form produced by strand transfer into two discrete molecules. The CCD in isolation is capable of conducting this reaction. (Adapted from [155])


A.
3'-Processing
5'
3'
CAGT
GTCA
5'
3'
CA
GTCA
T-
Mg2+
Mn2+
Substrate
Product
B.
Strand Transfer
5'
3'
CA
GTCA
3'
5'
GTCA
5'
CA
T-
Mg2+
Mn2+
Strand
Tranfer
Products
Substrate
C.
Disintegration
T-
Mg2+
Mn2+
Substrate
GT
CA
Product

3' Processing-

3' processing (3' P) is the initial step of the integration reaction, and is presumed to occur in the cytoplasm, concomitant with the completion of the reverse transcription program. The reaction proceeds initially with IN recognition and binding of the viral DNA substrate, which is facilitated by *cis*-acting attachment sites (*att*) found at the extreme termini of the DNA substrate [156]. The nucleotides encompassing the *att* constitute approximately 10 nucleotides, and contain a highly-conserved CA dinucleotide that is absolutely essential for the integration reaction (Figure 9). *In vitro* mutagenic analysis of the *att* region demonstrates the absolute necessity of the CA nucleotide for the integration reaction, and a less stringent requirement of the sub-terminal nucleotides [135]. After attachment, IN proceeds to cleave the terminal –GT dinucleotide from the 3' strand at both the U3 and U5 regions, just after the conserved –CA dinucleotide. This is accomplished by nucleophilic attack of the phosphodiester bond by a water molecule. The recessed 3' strand thus produced possesses a free –OH group that is used as the nucleophile in the subsequent strand transfer reaction.

Though high resolution images of the IN-DNA complex are non-existent, several lines of evidence indicate that 3' P is conducted by the action of a dimer of IN molecules acting at each end of the viral DNA. These include observations of the efficient processing activity of cross-linked dimers versus monomeric and tetrameric forms [157], and the concurrent loss of processing activity in conditions in which IN tetramerization is favored [158, 159] . HIV-1 IN displays a natural propensity to multimerize and to form

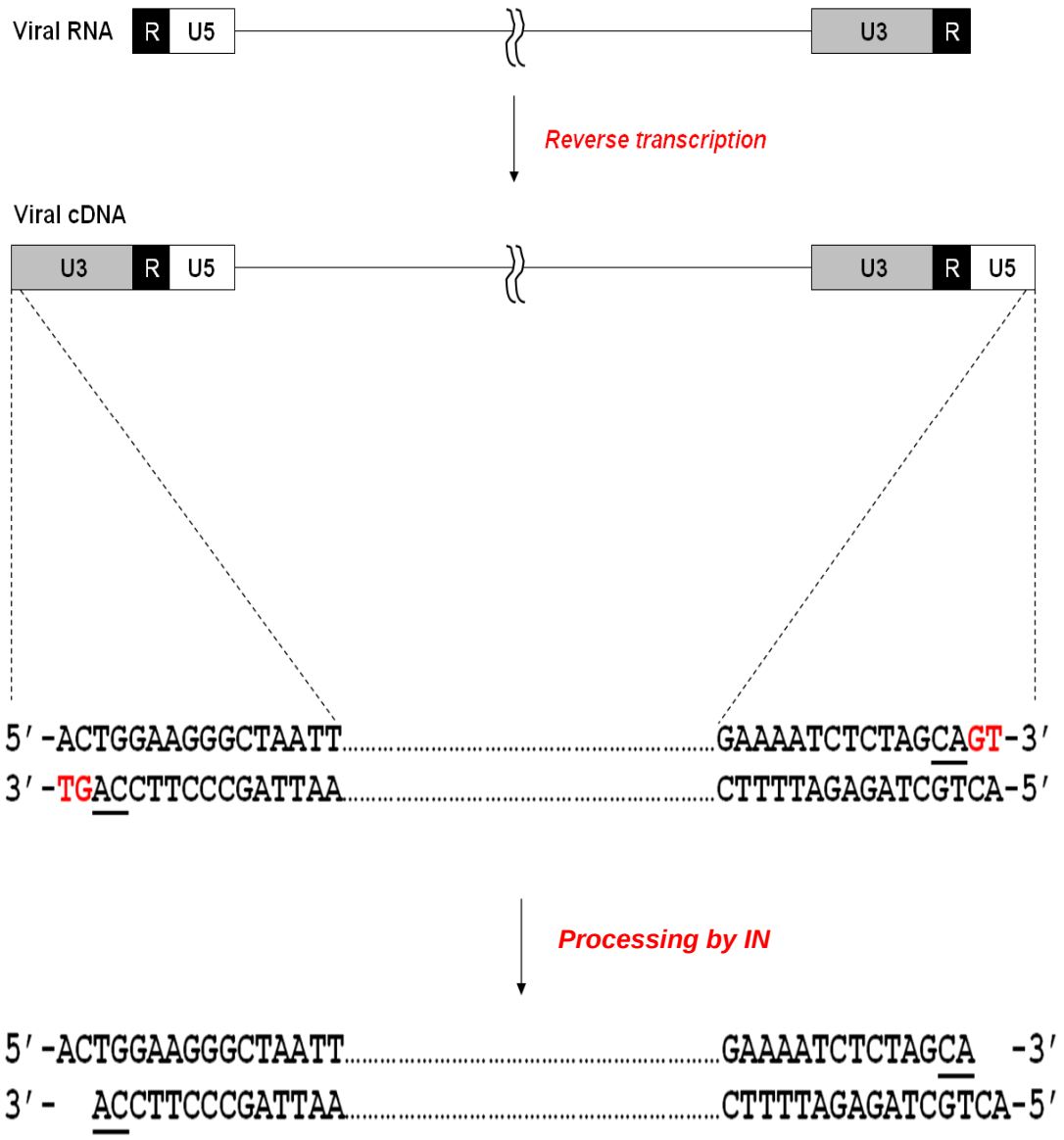


Figure 9. The 3' processing reaction. The reverse transcription process generates the LTRs that flank the viral cDNA. At the termini of these ends are attachment sites (*att*) that facilitate IN binding and processing. The CA dinucleotide, shown underlined, is a highly conserved sequence found in retroviruses, and is essential for integration activity. HIV-1 IN removes the –GT dinucleotide from both end of the viral DNA, activating them for the subsequent strand transfer reaction.

aggregates; in solution a dynamic equilibrium of monomers, dimers, tetramers, and other high-order complexes of IN is established [142, 160]. The importance of multimerization for integration activity is reflected in the lack of catalytic activity of monomeric IN [157], and the abrogation of integration activity when residues known to affect IN multimerization are mutated [131, 161]. The current model for the IN-DNA complex during 3' P is shown in Figure 10.

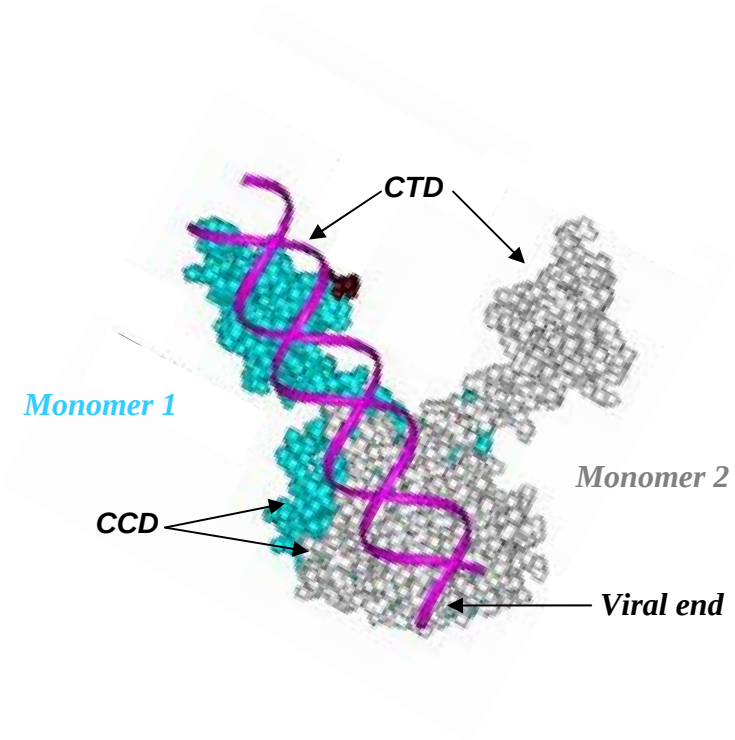


Figure 10. IN dimer model for 3' processing. This model of IN action for processing and integration activity is based on observations from complementation studies, crosslinking analyses, and structural data. The CTD of monomer 1 in the background associates with subterminal sequences in the viral LTR, orienting it such that the viral end makes contact with, and is processed by the catalytic core of monomer 2 in the foreground. Shown in maroon is Glu 246, a CTD residue found to crosslink to viral DNA. (Adapted from [132]).

This theoretical model closely resembles the resolved two-domain crystal structure of the CCD and CTD, consisting of two IN molecules associated exclusively via CCD interactions, with the CTDs of the two subunits splayed apart by 55 angstroms, forming a Y-shaped dimer [124]. The proposed orientation of the viral DNA substrate with the IN dimer is such that the CTD of one monomer stabilizes the DNA so that the DNA end lies in the catalytic site of the opposed dimer. This model satisfies certain observations of IN biology. For example, it is well-established that two unique mutant IN monomers can very efficiently complement each for catalysis if the mutations lie in distinct domains [162]. Additional support for the IN-DNA arrangement during processing is derived from the results of several crosslinking studies, demonstrating the proximity of catalytic core residues to nucleotides at the extreme terminus of the viral DNA, while subterminal nucleotides interact specifically with residues in the CTD [163-166]. During or after processing of the viral ends, it is believed that IN undergoes a conformational change that engages the complex for the subsequent strand transfer reaction [132, 163, 166].

Strand transfer-

The strand transfer reaction is contingent on the previous IN activity of viral DNA end processing. The removal of the dinucleotides generates reactive –OH groups at the 3' ends of the viral DNA, which then IN directs in nucleophilic attack of phosphodiester bonds in the target DNA. The cleavages occur along one face of the target DNA molecule, in the characteristic 5 bp staggered cut of HIV-1 IN. The cleavage of target DNA and the ligation of the activated viral DNA ends occur in concert, the resulting

product consisting of viral DNA attached covalently to host DNA at both 3' ends. The staggered cutting of target DNA results in a gap of single-stranded DNA 5bp in length at both ends of the viral DNA. The 2bp overhang of viral DNA created during 3' processing remains unhybridized. Host DNA repair enzymes then repair the gaps, completing the formation of the provirus. The repair of the single-stranded regions flanking the integration site results in the duplication of a 5bp sequence that is characteristic of HIV-1 IN integration sites. (Figure 11).

Current theories posit that the strand transfer reaction is conducted by at least a tetrameric assemblage of IN subunits, created by association of the two IN dimers that had functioned at either DNA end during the previous 3' processing reaction. Evidence for this stems from *in vitro* work demonstrating the IN tetrameric form is responsible for concerted integration reactions, whereas dimers can only facilitate the half-site reaction [157]. Furthermore, it was found that LEDGF binds specifically to tetrameric IN, and enhances concerted integration efficiency [167, 168]. A model for the assembly of IN with substrate and target DNA is shown in Figure 12.

Disintegration-

Observed exclusively *in vitro*, this biochemical activity of IN is essentially the reversal of the strand transfer reaction. A branched DNA substrate, such as shown in Figure 8, resembling the immediate product of strand transfer, is resolved by IN into two discrete molecules. The CCD in isolation is capable of mediating this reaction [169], underscoring the fact that the CCD houses the primary catalytic activity of IN. The

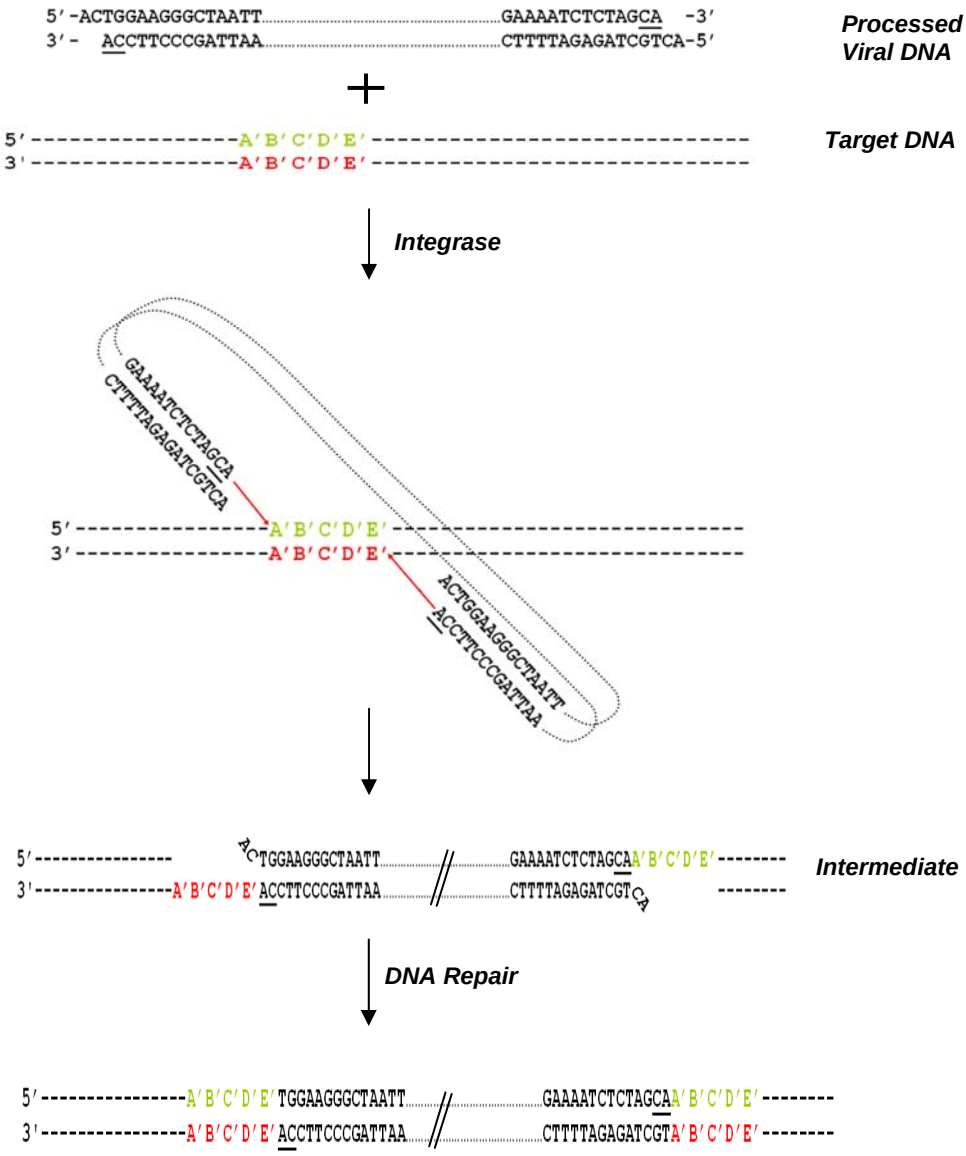


Figure 11. The strand transfer reaction. The substrate for the strand transfer reaction is the product of the 3' P reaction that immediately precedes it. The exposed –OH groups on the processed viral DNA ends are used in a nucleophilic attack of phosphodiester bonds in the target DNA, resulting in the covalent linkage of the recessed viral DNA end with target DNA. The 5 bp staggered cut ultimately results in the duplication of this original target DNA sequence at either end of the provirus.

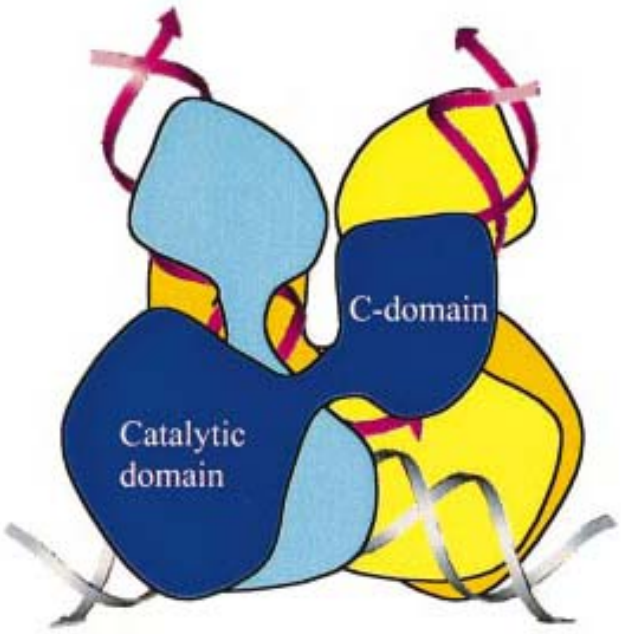


Figure 12. IN tetramer model for strand transfer. An IN two-domain tetramer model, assembled using constraints from crosslinking and structural data, is depicted here with associated DNAs. The two viral LTRs are colored in pink, and target DNA colored in gray. The light and dark blue monomers represent one dimeric unit attached to one viral DNA end, which associates with the yellow/orange dimer with its associated DNA. (Adapted from [163])

biological relevance of this activity is uncertain, and may simply be an artifact of the *in vitro* system in which it is studied.

Secondary Integrase functions:

In contrast to the primary integration activities, far less is known of the secondary functions of IN. It was long observed that gross deletions of IN negatively impact the yield of virus particles from transfected cells. Other IN mutants, which appear otherwise normal, were found to be blocked at a stage after viral release, but prior to integration. The Class I/Class II scheme formulated for convenient categorization of IN mutant phenotypes is little more than a broadly descriptive designation, especially so for Class II mutants that display extensive pleiotropic effects on viral dynamics. Advantageously, new mutants can readily be classified using this scheme, by measuring the levels of late RT cDNA product in newly infected cells; most Class II mutants, despite the nature of the block, are significantly less efficient at producing viral cDNA compared to Class I or WT viruses. The only exceptions are mutants that may be blocked for nuclear import, but not for integration itself.

Further investigation into the phenomenon of the IN-mediated effect on viral release revealed that the disruption is based on aberrant processing of the viral polyprotein precursors within the producer cell. Whereas processing should occur during or after virion egress from the plasma membrane, it was apparent that deletion and some point IN mutant viruses were processing the viral polyproteins intracellularly [47]. The timing of protease activation after completion of assembly is crucial for the proper packaging of viral components into the virion. Insufficient packaging of the Pol products

into outgoing virions can also have downstream consequences for virion morphology, reverse transcription and integration. Prematurely active PR can also have a direct detrimental impact on assembly, since the process itself is dependent upon the contributed activities of several domains of the intact Gag precursor. The precise mechanisms that avert the premature activation of the viral protease within the cytoplasm are currently unknown. Dimerization of Gag-Pol precursors is thought to be the rate-limiting step; once it occurs, the embedded viral protease self-catalyzes its own excision by an intramolecular mechanism, and then proceeds to cleave adjacent polyproteins in an ordered fashion [49]. The cytoplasmic ratio of 20-Gag to 1-Gag-Pol precursors in the producer cell is likely a natural deterrent to Gag-Pol dimerization. Indeed, artificially overexpressing Gag-Pol leads to premature PR activation in the cytoplasmic compartment [36, 170]. Furthermore, when added to virus producing cells, the non-nucleoside RT inhibitor (NNRTI) efarivenz, which enhances the dimerization of the subunits of the RT enzyme, triggers the premature activation of the viral protease [171].. Speculatively, IN in the context of the Gag-Pol precursor must exert some influence over the dimerization potential of Gag-Pol, perhaps by preventing a productive association that would inexorably lead to PR activation.

Significant headway has been made in recent years to unravel the role that IN plays in the reverse transcription process. The complicating factor has always been that many Class II IN mutants can be defective at steps prior to the early life cycle events which encompass reverse transcription. The development of an elegant method of targeting RT and IN *in trans* to virus particles independently of Gag-Pol [172], has

effectively created a system in which viral replication dynamics can be studied in the absence of IN influence on late life cycle events, and has thus opened the way for investigating the specific impact of IN on reverse transcription. Using this technique, it was clearly shown that mature IN is required for the production of viral DNA in infected cells, via specific interactions with components of the reverse transcription complex [108, 109]. Subsequent studies revealed a physical association between RT and IN, which is mediated by the IN CTD [110, 149]. However, the components and composition of the PIC are unknown, and it is also feasible that IN could play less direct role in the reverse transcription process, perhaps by recruiting other factors that interact with RT, or by contributing a scaffolding function to foster an appropriate environment in which reverse transcription can proceed.

IN has also been demonstrated to selectively recruit cellular proteins into virions [173, 174]. One of these, the host DNA repair enzyme uracil DNA glycosylase 2 (UNG2) has been shown in other studies to be incorporated by Vpr [96]. The enzymatic activity of UNG2 removes uracil residues from G:U mismatches in double-stranded DNA, and its recruitment to virions correlates with an enhancement of the fidelity of reverse transcriptase [96]. This activity appears to be particularly pertinent for infection of macrophages [175], which contain inherently low cytoplasmic dNTP levels, and high uracil levels [176]. Another cellular protein that was found to be recruited by IN into virions is INI1/hSNF5, a component of the chromatin remodeling SNF-SWI complex [177]. Though it is packaged into virions, this protein's effect on viral replication is at the stage of viral assembly, impacting the efficiency of viral release [174]. Given the timing

of this effect, it is apparent that INI1/hSNF5 interacts with IN in the context of Gag-Pol, and influences events in the late stage of the viral replication program. A follow-up to this study revealed that both INI1/hSNF5 and IN interact with SAP18, a component of the Sin3a-HDAC1 complex, and that components of this complex are incorporated into virions [178]. In keeping with this recruitment, HIV-1 virion preparations were subsequently demonstrated to possess a significant level of deacetylase activity, which if removed by overexpression and incorporation of a transdominant negative mutant of HDAC1 negatively impacted the early reverse transcription stage. This observation is especially intriguing, given that IN is modified by acetylation, and is involved in the efficiency of reverse transcription.

CHAPTER 2- MATERIALS AND METHODS

Construction of integrase mutants. The HXB2-related virus, R7/3(X/S) [4, 179], was used as the reference, wild type virus for the studies presented. Its cognate proviral molecular clone, plasmid pR7/3(X/S) (Figure 1), contains an XbaI restriction site at the 5' end of the integrase coding region and a SacII restriction site near its 3' end, the sites silent with respect to the overlapping viral coding frames and allowing for a cloning strategy that provides unbiased comparisons to be made against an otherwise isogenic background. Thus, using overlap PCR mutagenesis [pR7/3(X/S) template; outside oligonucleotide primers specifying XbaI and SacII sites], a series of translational nonsense mutations (amber) were introduced at sequential codon positions to create a nested set of single amino acid deletions originating from the IN C-terminus; all mutant proviral clones verified by DNA sequence analysis of the across the entire XbaI/SacII recombinant segment. Wild type and mutant XbaI/SacII inserts were also reconstructed into pR7/3(X/S)Bsd [4, 179], an *env*- proviral derivative in which the amino terminus of *nef* is substituted with *bsd*, a gene conferring resistance to the microbial antibiotic, blasticidin S. Single-step infectious, VSV-G pseudotyped viral stocks were prepared by co-transfection of the pR7/3(X/S)Bsd variants with pCI-VSV-G, a vesicular stomatitis virus glycoprotein (VSV-G) expression vector.

Tissue culture and viral stock preparation. HEK293T, GHOST(3)X4/R5[180] and HeLaP4R5 [181] cells were cultured in Dulbecco modified Eagle medium (DMEM)

supplemented with 10% fetal calf serum (FCS) and 100 U/ml penicillin/100 µg/ml streptomycin (P/S); CEM cells were cultured in RPMI 1640 supplemented with 10% FCS and P/S. Viral stocks were generated in one of two ways. For multicycle, replication-competent (*env*+) virus, stocks were made by mixing 2 µg of proviral plasmid DNA with 12 µg polyethylenimine (PEI) and adding the mixture to a sub-confluent monolayer of HEK293T cells cultured in 6-well plate format. Single-step viruses were made by mixing 1.5 µg viral plasmid DNA and 0.5 µg pCI-VSV-G expression vector with 12 µg PEI and adding to HEK293T cells. For both methods, media was changed the following day and the viral stocks harvested 48 hours post-infection (hpi) by filtration of cell-free media through a 0.45 µm nylon filter. Viral titer was determined using data provided by p24 ELISA (PerkinElmer Life and Analytical Sciences).

Western blot analyses. ***Intracellular protein*** Transfected HEK293Tcells were disrupted by lysis in RIPA buffer (10 mM Tris [pH 7.5], 150 mM NaCl, 1% deoxycholate, 1% Triton X-100, 0.1% sodium dodecyl sulfate [SDS]) supplemented with "cOmplete" protease inhibitor cocktail (Roche) and the lysates cleared by centrifugation at 20,817 x g for 15 min at 4˚C.

Virion protein Viral stocks, generated by proviral DNA transfection of HEK293T cells, were normalized for p24 content (250 ng) and pelleted through a 25% sucrose cushion at 20,817 x g for 3 hours at 4˚C and the pelleted virions lysed in supplemented RIPA buffer. Intracellular and virion protein samples were normalized for protein content [182] and 25 µg (intracellular) or 15 µg (virion) loaded onto 4-12% polyacrylamide NuPage Bis-Tris

gels (Novex) and using reducing conditions, resolved by electrophoresis in MOPS buffer. Proteins were blotted onto polyvinylidene difluoride (PVDF) membranes, which were subsequently blocked, washed and incubated with monoclonal antibodies directed against the amino terminus of IN (6G5) [183], capsid (CA) (183-H12-5C) [184] or RT (11B7 and 33D5) [185]. Blots were then incubated with rabbit anti-mouse mAb-HRP, and detected using the Immobilon Chemiluminescent AP Substrate Detection Kit (Millipore).

Qualitative Assay for Viral Replication Competency. Detection of syncytia formation in the highly fusogenic CEM cell line after infection is indicative of successful proviral DNA integration and efficient *env* gene expression—the rate and extent of syncytia recruitment and expansion over time, a qualitative measure of viral replication-competency or fitness over several cycles of growth. Wild type or mutant virus stocks were normalized for equal p24 content and used to infect CEM cells (10^3) with a viral inoculum of 5 ng p24 antigen in 200 µl of medium in 96-well round bottom plate format or mock infected with medium alone. All infections were performed in duplicate. The cells were microscopically monitored daily for syncytia formation over the span of a 10-12 day observational period. Seven days post-infection (7 dpi) the cultures were split 1:40 and again examined daily for an additional 7 days.

Genetic Integration Assay (*bsd* Transduction). HeLaP4R5 cells seeded to approximately 30% (24-well plate format) and inoculated overnight with a normalized amount of virus (5 ng p24). Twenty-four hours later the cells were washed twice with

PBS and fed with fresh DMEM for an additional 24 hour growth period, after which the cells were split into selection media containing 5 μg/ml blasticidin HCl. Selection was imposed for an additional 14 days and then resistant colonies were fixed, stained and scored macroscopically for the number of resistant colonies.

Quantitative PCR Analyses. ***Late RT and 2-LTR*** Viral stocks were pretreated with Turbo DNase (Ambion) at a concentration of 40 U/ml at 37˚C for one hour before infecting confluent GHOST(3)X4/R5 cells with 50 ng p24 per well in 24-well plate format in the presence of 8 μg/ml polybrene. Infections were spinoculated for 2 hours at 15˚C at 2400 rpm 1,160 x g before transferring the cultures to a 37˚C humidified incubator for 8 hours. At this time, total cellular DNA was prepared for late RT product determinations, or 16 hours later (24 hpi) for the quantification of 2-LTR circular DNA accumulation. The QiaAmp DNA Blood Mini Kit (QIAGEN) was used for all isolations, the extent of late RT and 2-LTR circles formation determined by molecular beacon-mediated quantitative PCR analysis using primers specific for each type of amplification. Late RT: Forward primer (5'-AGATCCCTCAGACCCTTTTAGTCAGTGTGG-3') with reverse primer (5'-GCCGCCCCTCGCCTCTTG-3') and beacon (5'-\56-FAM\ccgacccTCTCGACGCAGGACTCGGCTTgggtcgg\3Dab\-3'); 2-LTR Circles: Forward (5'-CTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTA-3') with reverse primer (5'-TGACCCCTGGCCCTGGTGTGTAG-3') and beacon (5'-\56-FAM\ccgcacC TACCACACACAAGGCTACTTCgtgcgg\3Dab\-3').

Alu PCR Integration Assay Infections and DNA extractions were performed identically as for 2-LTR quantitative PCR analyses. Using a modification of the twin PCR amplification protocol designed by Chun et al. [186], extracts were normalized for DNA content and then subjected to two rounds of PCR, the first of which uses a forward primer specific to the Alu repetitive element (5′-TCCCAGCTACTCGGGAGGCTGAGG-3′) and a reverse primer (5′-AGGCAAGCTTTATTGAGGCTTAAGC-3′) localized to the U3 region of the HIV-1 LTR. The reaction conditions for the first round PCR were as follows: 94˚C x 3 min and then 22 cycles of 30 sec x 94˚C, 30 sec x 66˚C with a 5 min extension reaction at 72˚C followed by a final 10 minute extension at 72°C. One/one hundred and twentieth ($1/120^{th}$) of the first round reaction was then subjected to a second round of PCR, here modified for beacon-mediated quantitative PCR analysis using a nested primers (5'-GAAGGGCTAATTCACTCCCA-3', 5'-CTTGAAGTACTCCGGAT GCAG) in the LTR in conjunction with a molecular beacon (5'-/56-FAM/ccgcacCTACCACACACAAGGCTACTTCgtgcgg/3DAB/-3'). The second round PCR reaction used the following reaction conditions: 10 min x 94˚C and then 50 cycles of 30 sec x 94˚C, 33 sec x 63˚C and 30 sec x 72˚C. Integration standards for qPCR analysis were produced by infecting HeLa P4R5 or GHOST(3)X4/R5 cells at a low multiplicity of infection (moi) of BSD (WT IN) virus, subjecting cells to a week-long selection at low density, and extracting total DNA 8 dpi. DNA from these cells was subjected to the first round PCR, and diluted 40x. A dilution series of this 40x diluted first round PCR was used as the standard by which the tested viruses were compared for quantification.

Ligation-Mediated [187], Quantitative PCR (LMqPCR) Assay for 3´ Processing

Infections and DNA extractions (8 hpi) were carried out under identical conditions as that used for late RT PCR analyses. Here, extracts were normalized for DNA content and then incubated for 12-16 hours at 16˚C in a 15 µl reaction mixture containing 500-750 ng total cellular DNA and an equimolar amount of two oligonucleotide primers (each 2 µM final): *LMqPCR-short* (5´-OH-GTACTCATGTA-OH-3'), *LMqPCR-long* (5´-OH-GTCTAGAGCTCAGCTGTACATGAGT-OH-3'). LMqPCR-short (11 base-pairs) is complementary to LMqPCR-long over 9 contiguous nucleotides (LMqPCR-short nucleotides 3-11) such that when annealed, the double-stranded oligonucleotide produces a two base-pair 5' overhang (nucleotides G_1 and T_2), the exact compliment to the 2 base-pairs left unpaired after integrase-mediated 3' processing at each end of proviral DNA (5'-AC-3'), the rest of LMqPCR-long remaining single-stranded and used to anchor a subsequent quantitative PCR assay. Upon ligation, the single phosphate group at each of the 5' ends of linear proviral DNA is covalently linked to the 3' end of the LMqPCR-long primer DNA. Ligation reactions were performed with 10 U of *E. coli* ligase (Takara) in 1x *E. coli* DNA ligase buffer (New England Biolabs). Since both LMqPCR-short and LMqPCR-long are unphosphorylated, the oligonucleotide pair is relatively inert with respect to promiscuous ligation events, incapable of their individual oligomerization or the formation of multiple concatemers upon the ends of the proviral DNA. After ligation, the reaction was incubated at 65˚C x 20 min to inactivate the enzymatic activity of the ligase and then one-third of the ligation reaction (5 µl) was supplemented with the addition of the LMqPCR-long primer and subjected to qPCR analysis, the ligated DNA

products quantified using beacon-mediated qPCR analysis using LMPCR-long and an HIV-1 LTR primer/beacon oligonucleotide set specific for monitoring the efficiency of 3' processing events at either the 5' or 3' LTR. 5' LTR primer/beacon pair: (5'-CTTGCTCAACTGGTACTAGCTTGTAG-3') used in conjunction with (5'-/56-FAM/ccgcacCTACCACACACAAGGCTACTTCgtgcg g/3DAB/-3'); 3' LTR primer/beacon pair: (5'-GGGAGCTCTCTGGCTAACTAGG-3') used with (5'-/56-FAM/ ccgaaccaGTAGTGTGTGCCCGTCTGTTGTGtggttcgg/3DAB/-3'). Controls for all quantitative PCR reactions above include IN D116A virus (no IN enzymatic activity) and an RT mutant virus (D185A/D186A) [4] with no polymerization activity, the latter a control for bacterial contaminant DNA left over from the preparation of viral stocks. An additional ligation specificity control was included for LMqPCR analyses. Here, an oligonucleotide (LMqPCR-short$^{\mathbf{TG}}$, 5′-OH-TGACTCATGTA-OH-3') complementary to LMqPCR-long and identical to LMqPCR-short except for the first two nucleotides where G_1 and T_2 have been reversed (T_1 and G_2). As a result of this modification, LMqPCR-short$^{\mathbf{TG}}$ is mismatched with the unpaired 5'-AC-3' dinucleotide left after integrase-mediated 3' processing, the LMqPCR-long/ LMqPCR-short$^{\mathbf{TG}}$ oligonucleotide pair unable to be efficiently ligated during the first step of the procedure.

Southern blotting. Viral stocks were pretreated with Turbo DNase (Ambion) at a concentration of 40 U/ml at 37˚C for one hour before infecting confluent GHOST(3)X4/R5 cells with 500 ng p24 per well in 6-well plate format in the presence of 8 μg/ml polybrene. Infections were spinoculated for 1 hour at 25˚C at 2400 rpm (1,160 x

g) before transferring the cultures to a 37˚C humidified incubator for 26 hours. At this point, total cellular DNA was extracted using the QiaAmp DNA Blood Mini Kit (QIAGEN). 5 µg of DNA were digested with 20U of BamH1 (NEB) and 9U of Msc1 (NEB) overnight at 37°C, heat-killed and electrophoresed on a 0.6% agarose gel, and then transferred to nitrocellulose membrane by upward capillary action. The transferred DNA was crosslinked to the membrane using a Strategene UV-crosslinker, and the membrane was incubated for 10 minutes in a prehybridization step with SIGMA Perfect Hyb (Catalogue# H7033) containing 100 µg/mL denatured salmon sperm DNA at 68°C in a rotation hybridization chamber. The probe was synthesized by PCR amplification of a region of the R7/3 (X/S) plasmid flanking the upstream Msc1 restriction site (Primers: 5'-CTAGAAGAAATGATGACAGCATG-3'and 5'-CTGTTAGTGCTTTGGTTCCTCT AAG-3'). The PCR product was gel purified and radioactively-labeled with α-P^{32}-ATP (Perkin Elmer Easytides nucleotides 100µCi, Cat# BLU512H100UC) by random priming using the Ambion Decaprime II kit (Cat#1455). Unincorporated nucleotides were removed by passing the reaction over a NucAway Spin column (Cat# AM10070). The probe was denatured for 10 minutes at 100°C, incubated on ice for two minutes, and added to the prehybridization solution with the membrane, which was incubated with rotation at 68°C overnight. The membrane was initially washed in 2x SSC in 0.1% SDS wash buffer at room temperature. Subsequent washes were carried out at 68°C with 0.5x SSC/0.1% SDS and 0.2x SSC/0.1% SDS for 15 minutes each. The blot was then exposed to a phosphor screen, which was developed on a Storm Imager. Image analysis was conducted with the supplied software.

Microscopy and IN chromatin binding studies. 293T cells plated at 30% confluency on poly-L-lysine-treated (0.01% solution, Sigma) coverslips were transfected with 0.5 μg plasmid DNA encoding various eGFP-IN truncation fusion proteins. At 12 hours post-transfection, cells were washed, fixed with 4% paraformaldehyde and mounted onto microscope slides with Vecta-Shield mounting media containing 4,6-diamino-2-phenylindole (DAPI) for nuclear staining (Vector Laboratories, Burlingame, CA). An inverted DeltaVision optical sectioning microscope (Applied Precision) was used to collect between 70 to 80 images spaced at 0.2 μM through the slide. The images were deconvolved using the DeltaVision deconvolution software (Applied Precision).

Primary cells culture. Primary memory CD4+ T were isolated by magnetic negative selection from Ficoll-Paque (Catalogue# 17-1440-02) purified peripheral blood mononuclear cells (PBMC) over an LS column as per the manufacturer's protocol (Miltenyi Biotech, Catalogue# 130-091-893). The purified cells were cultured in RPMI 1640 supplemented with 2% heat-inactivated human serum and 100 U/ml penicillin/100 μg/ml streptomycin (P/S). Cells were stimulated with 5 μg/ml phytohemagglutinin (PHA) for approximately 12 hours prior to viral challenge.

RNA extraction and Reverse transcription. Transfected or infected cell lysates were homogenized on a QIAGEN QIAshredder homogenizer, and total RNA was retrieved from the cleared lysate using the QIAGEN RNeasy Mini kit. Any contaminating genomic

or viral DNA was removed by on-column digestion with the RNase-Free DNase Set (Catalogue# 79254). The RNA was reverse transcribed using the Thermo X reverse transcription kit (Invitrogen, catalogue# 11150100) using recommended parameters as per the manufacturer's directions, and the cDNA was subject to amplification using a primer upstream of the HIV-1 major splice donor (MSD), and another specific to the Vif or Vpr spliced transcripts.

Viral evolution experiments. A replication-competent IN 269 (5x) stop virus was used in the viral evolution experiments. This virus contains five engineered stop codons: three in the IN reading frame, and two others in the two alternative reading frames downstream of the 269 codon (Figure 33). Two virus backgrounds were used: Vpr-, vpu-, and vpr-vpu+. CEM-SS cells were infected with the viruses, and passaged every four to five days, or when cells became confluent, for several weeks. After every three or four cell passages, viral supernatant from the infected cultures were spun down to pellet the cells, and the cell-free media used to commence infection of fresh cells. Cultures were monitored for the formation of syncytia every few days. Concurrent with the observation of an unusual rate of syncytia formation in the cell culture, total DNA was extracted from cells, the IN region of the virus was amplified by PCR, and sent for sequence analysis of the bulk PCR product. The presence of revertant viruses was ascertained by observing the chromatogram traces. The PCR amplified viral gene products were subsequently cloned back into the parental plasmid, transformed into the DH10B *E. coli* strain, and individual

colonies were picked, grown up, and the viral DNA plasmid was sequenced to detect the presence of individual viral revertant clones.

CHAPTER 3- RESULTS

Preliminary Findings

REPLICATION COMPETETENCY OF IN TRUNCATED VIRUSES:

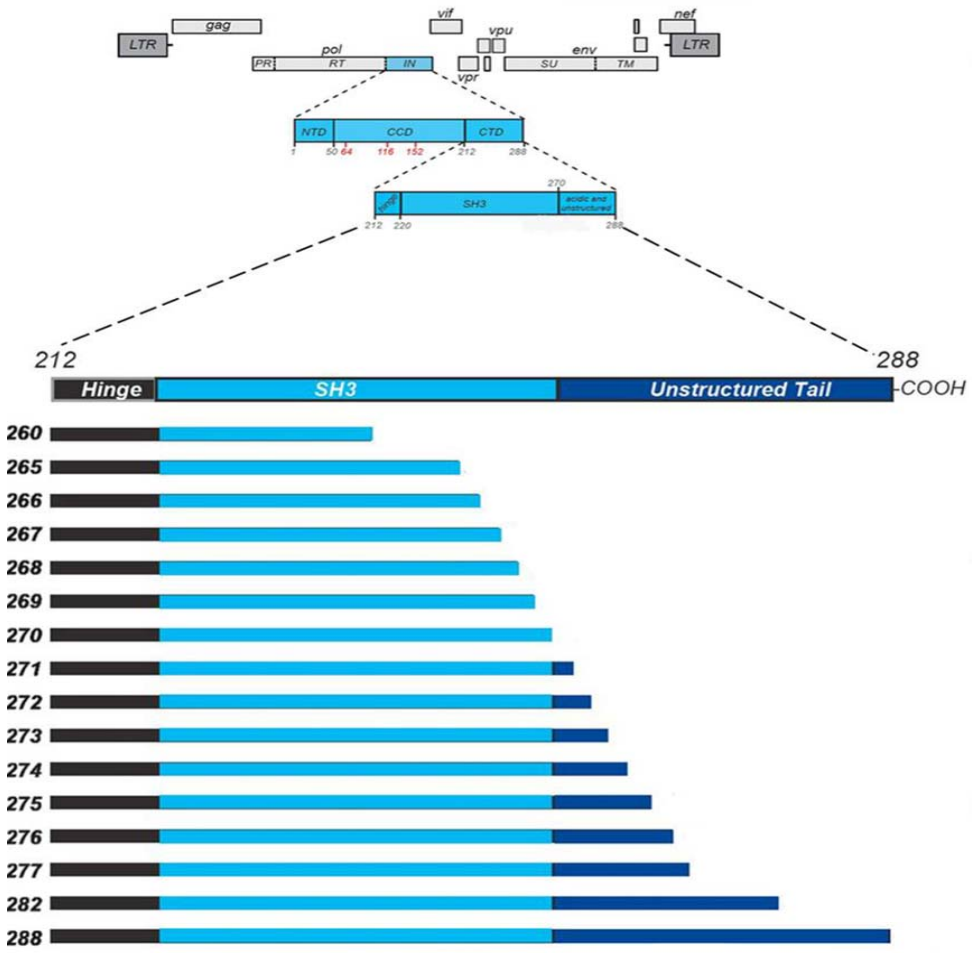
Microscopic examination of syncytia formation is a convenient method to monitor the extent of viral replication through cell culture *in vitro*, as both viral integration and proviral gene expression of the viral envelope glycoprotein (gp120) are required for its appearance. Using this assay, the panel of C-terminal integrase (Gag-Pol) truncation mutants, shown in Figure 13, was initially tested to determine which mutants were capable of sustained viral replication in the susceptible host human T cell line, CEM-SS. This particular CEM strain is Vif-permissive and chosen for the initial characterization of the deletion mutants since a portion of the integrase coding frame overlaps with the 5'end of the *vif* gene (integrase codons 271-288) (Figure 5). Therefore, any possible influence of the associated mutagenesis of the Vif protein might have among the panel of integrase mutants was removed from consideration and irrelevant to the interpretation of the results obtained.

Results-

Qualitatively, the rate of syncytia formation appears to be in positive correlation with increasing length of the integrase protein, the WT virus (288 amino acids) producing detectable syncytia within 3 days and rapidly leading to cell death of the entire culture within 7 days (Figure 14). Intriguingly, although viral replication competency is lost after removal of the terminal 16 amino acids of the protein (i.e., mutant IN 272), it is regained in mutant IN 269 but lost again for all further truncations N-terminal to this position. The same results were obtained across wide multiplicities of infection (data not shown).

Figure 13. Panel of IN truncation mutants. (A) The HIV-1 genome is depicted at top, the ORFs of the viral proteins depicted as discrete blocks. The IN protein, encoded initially as the C-terminus of the Pol protein, is blown up to illustrate the domain architecture of IN. The CTD is further magnified to reveal the structural features of the domain: the α-helical hinge attaching the CCD to the CTD is depicted in black, the SH3 folded region (220-270) in light blue, and the unstructured tail 18 amino acid in dark blue. (B) Amino acid sequence of the CTD. The letters are colored according to the scheme used to depict the structural elements of the CTD in (A).

A.


LTR
gag
pol
PR
RT
IN
vif
vpu
vpr
env
SU
TM
nef
LTR
NTD
CCD
CTD
SH3
212
288
Hinge
SH3
Unstructured Tail
COOH
260
265
266
267
268
269
270
271
272
273
274
275
276
277
282
288

B.


212
220
230
240
250
260
270
ELQKQITKIQNFRVYYRDSRNPLWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQM
280
288
AGDDCVASRQDED

A.

Mutant	Sustained Replication
ΔIN	ND
IN 212 stop	ND
IN 260 stop	-
IN 265 stop	-
IN 266 stop	-
IN 267 stop	-
IN 268 stop	-
IN 269 stop	+
IN 270 stop	-
IN 271 stop	-
IN 272 stop	-
IN 273 stop	+
IN 274 stop	+
IN 275 stop	+
IN 276 stop	+
IN 277 stop	+
IN 282 stop	+
WT	+
IN D116A	-
RT-	-

B.

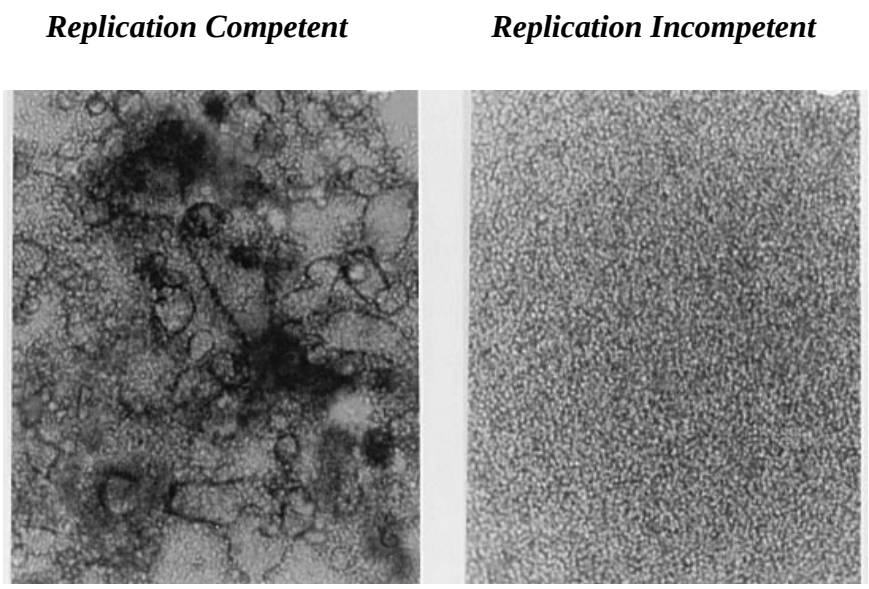


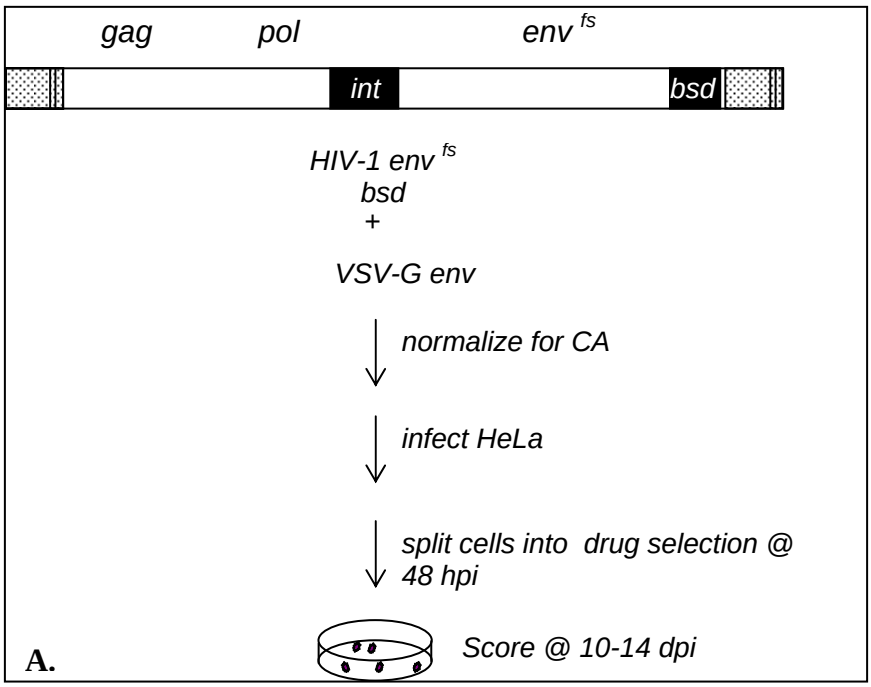
Figure 14. **Replication competency of IN truncations.** A) CEM cells infected with a low inoculum of replication competent IN truncated mutant viruses were observed daily for the formation of syncytia. + syncytia detected, - no syncytia observed two weeks post infection. ΔIN: a stop codon is introduced at the beginning of the IN ORF to remove the entire IN protein. RT-: control virus containing mutations in the reverse transcriptase protein that destroy RT catalytic activity, without affecting any viral function other than reverse transcription. ND: No Data. B) Replication competent viruses produce syncytia in CEM cells; cells infected with replication incompetent viruses continue to proliferate normally without detectable virus-induced cytopathic effects.

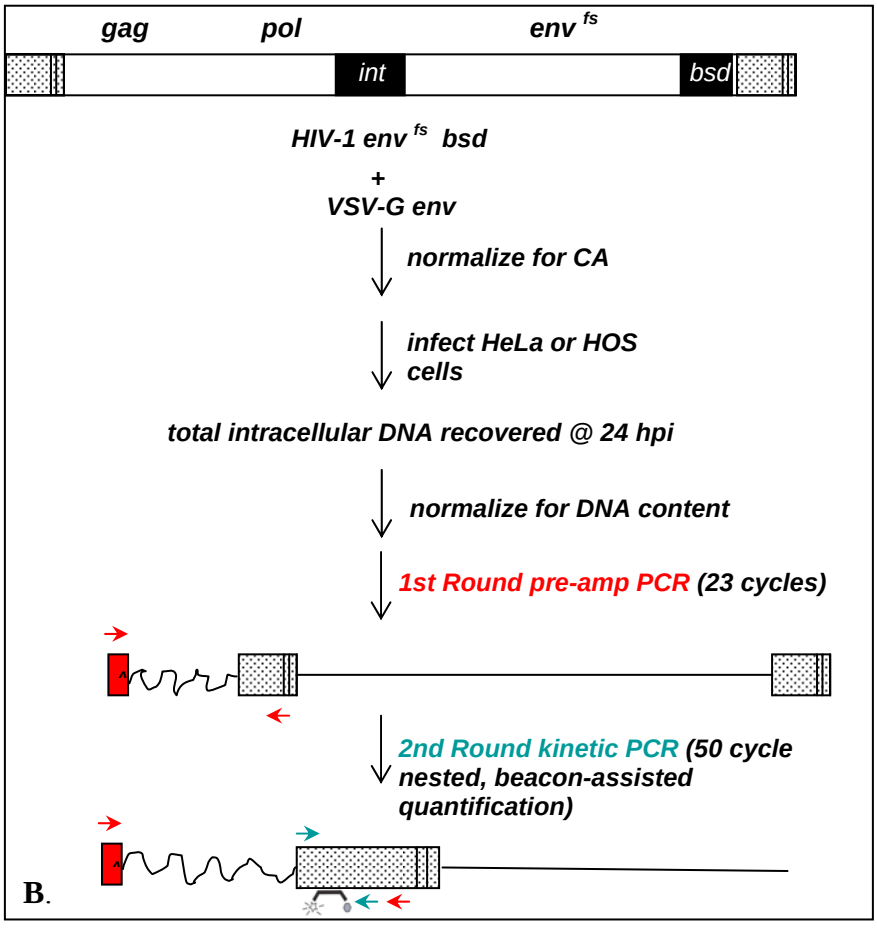
INTEGRATION FREQUENCY OF IN TRUNCATED VIRUSES:

It has been previously established that sustainable viral replication in culture is contingent on an integration rate that is at a minimum 10-15% of the wild type rate. Since alterations were being introduced into the enzyme responsible for integration, the integrative capacity of the truncated IN viruses were assessed. The obtained results could verify whether this previously established correlation of minimal integration rate and replication competency could account for the behavior of the IN truncation panel in the syncytia formation assay. The integration frequencies were measured in two distinct assays: one genetic, and one biochemical (Figure 15). First, in the BSD assay, we examined the ability of the viruses to stably transduce a susceptible cell line with a dominant selectable marker for blasticidin resistance. In order to specifically gauge integration and not viral replication dynamics, we limited infection to a single round by pseudotyping HIV-1 *env*$^-$ BSD viruses with the pantropic, vesicular stomatitis viral envelope glycoprotein (VSV-G). We included as an integration negative control the IN D116A catalytic mutant, a well-characterized Class I mutant known to possess virtually no integration activity, but efficiently performing all other integrase-facilitated processes. Blasticidin resistant colonies, each one of which is tantamount to a successful integration event, were scored, and for each virus, expressed as the percentage of the WT integration frequency (Table 2). The second method used to measure integration frequency is an adaptation of the Alu-PCR integration assay developed by Chun et al. Unlike the BSD integration assay, the readout of the Alu PCR assay is not contingent on viral gene

expression; instead, it is, by virtue of its qPCR component, a direct quantitative assessment of the level of the integrated form of viral DNA.

Figure 15. Assays for measurement of integration ability. (A) The BSD assay, integration is determined by macroscopic scoring of drug resistant colonies. B) The Alu-PCR Assay, a first PCR round amplifies DNA regions intervening integrated provirus DNA and Alu repeat elements. The second round uses nested primers to detect the amplified DNA from the first round.


gag
pol
env fs
int
bsd
HIV-1 env fs
bsd
+
VSV-G env
normalize for CA
infect HeLa
split cells into drug selection @ 48 hpi
Score @ 10-14 dpi
A.


gag
pol
env fs
int
bsd
HIV-1 env fs bsd
+
VSV-G env
normalize for CA
infect HeLa or HOS cells
total intracellular DNA recovered @ 24 hpi
normalize for DNA content
1st Round pre-amp PCR (23 cycles)
2nd Round kinetic PCR (50 cycle nested, beacon-assisted quantification)
B.

Results-

The integration rates of the IN truncation panel were measured by two distinct methods, the results presented in Figure 16 and 17. Side-by-side comparisons reveal the concordance of the independently-derived integration rates for each mutant virus. Furthermore, the values derived from both assays establish a similar trend of integration frequency across the panel of truncations. When examined in light of the observed syncytia-formation capabilities, the integration frequency trend correlates precisely both with an ability to replicate (≥10-15% of WT rate), and the rate with which syncytia overtook the individually infected cultures.

The discrepancy in the integration frequency of the IN D116A mutant in the two assays is due to an inherent drawback of the Alu PCR assay. IN D116A is a paradigmatic Class I mutant, producing copious levels of cDNA upon infection. Lacking integration catalytic activity, however, the DNA persists, at least in the short time frame of the Alu PCR assay (24 hours post infection) as linear and circularized DNA species. The first preamplifying PCR reaction of the assay serves to enrich Alu element-LTR intervening sequences, for detection in the second qPCR round by HIV-1 LTR nested primers. Though not specifically amplified in the first round, 1- and 2-LTR circular species yet present in the cellular DNA extract will nonetheless be carried over into the second round where they can function as a template for the HIV-1 specific nested primer pair. In this regard, the BSD assay is a more reliable measurement of integration, since unintegrated forms of DNA do not support levels of gene expression required to provide drug resistance. Furthermore, the long selection period (2 weeks) facilitates the loss of these

unintegrated species through dilution from multiple rounds of cell division, or by degradation.

Having established the respective integration capacity of each IN truncation mutant, and observing a correlation between these and replication competency, the next step was to determine why the integration rates were low in some mutants. Since mutations within IN can specifically affect the catalytic steps of integration, or alternatively, can induce otherwise pleiotropic effects, we initially endeavored to broadly classify our mutants according to the Class I/Class II scheme, establishing a stepping stone for more in-depth investigations.

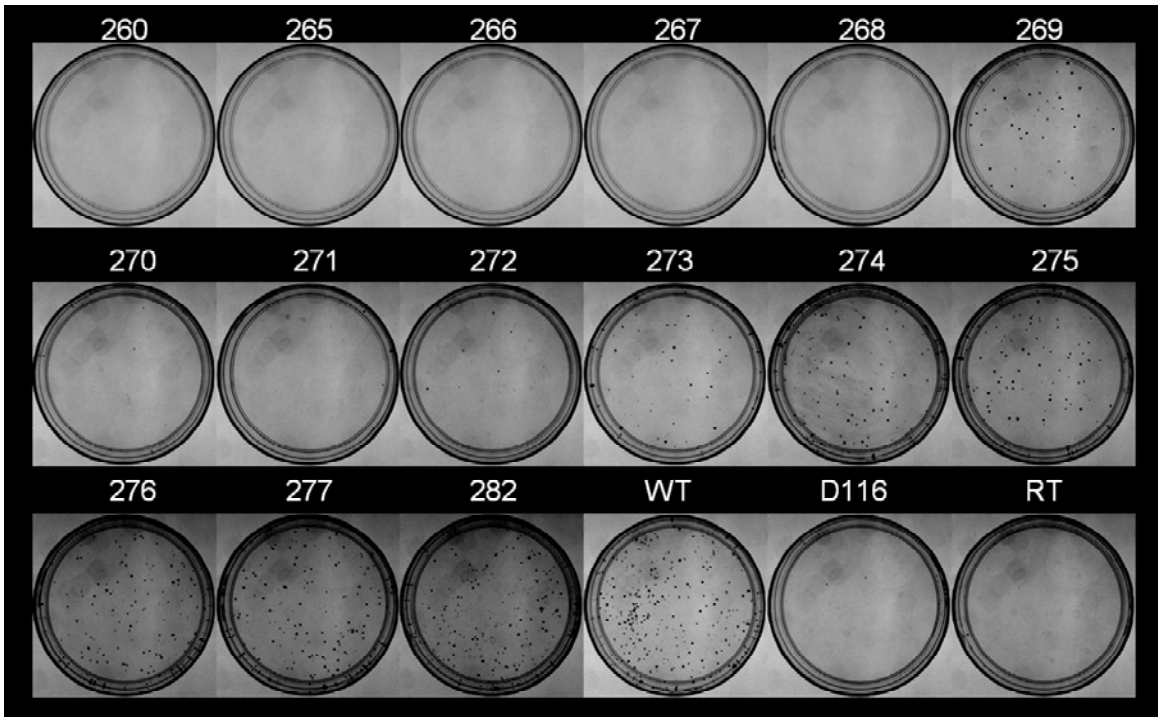


Figure 16. BSD Assay on IN truncations. Results are shown graphically in Figure 17.

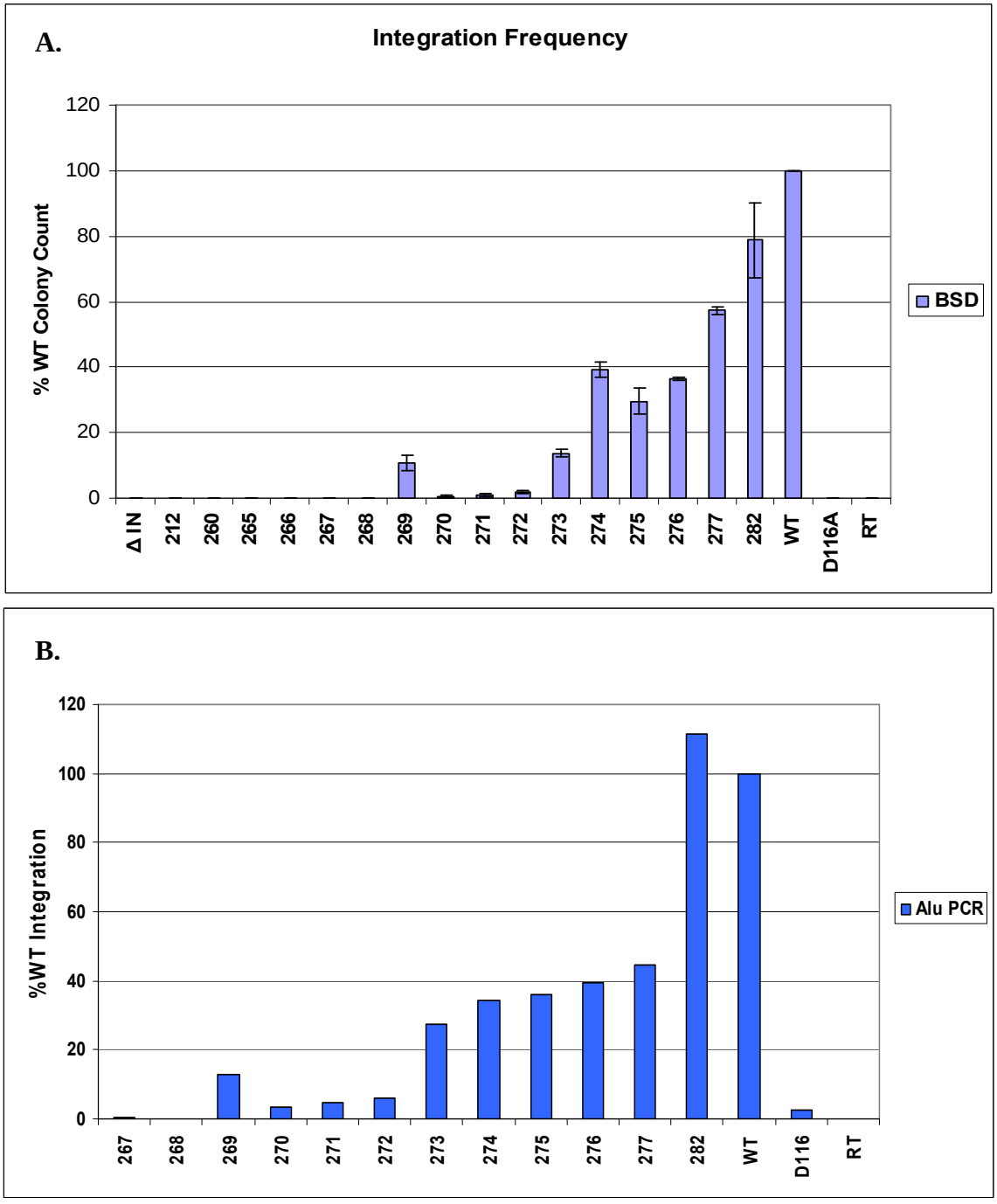


Figure 17. Integration frequency of IN truncations. The results of the (A) BSD and (B) Alu-PCR assay depicted graphically, the integration rates of the mutants expressed as a percentage of the WT rate.

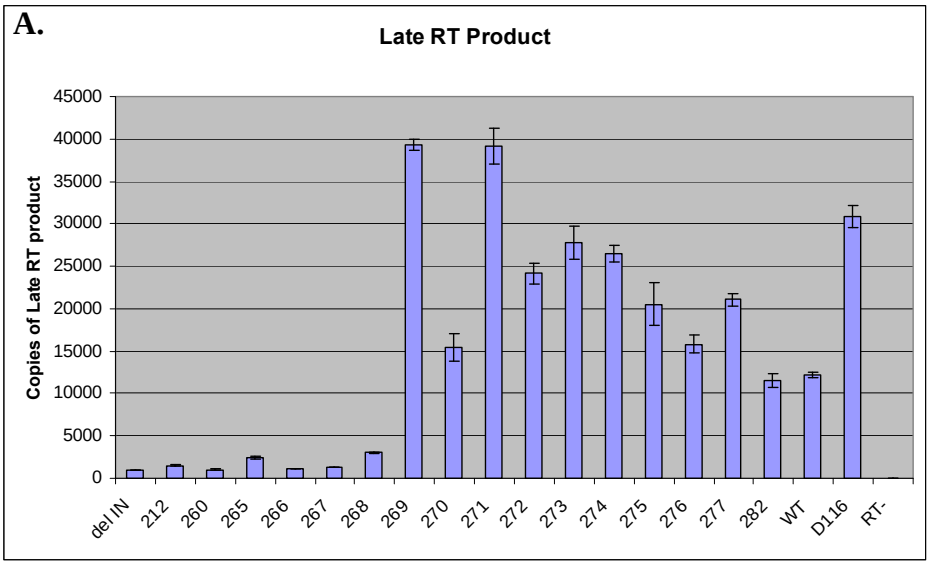
CATEGORIZATION OF NON-REPLICATING IN TRUNCATIONS INTO CLASS I/II PHENOTYPES:

The panel of IN truncation mutants can be segregated according their replication competency (Figure 14). We have established the integration/transduction efficiency of each and observed a direct correlation between the integration frequency of the respective truncation mutant and their ability to replicate through CEM cells. In some integrase mutants, the loss of transduction ability is rooted at the integration step itself. Alternatively, one or more points after proviral establishment could be affected including those steps (i.e., assembly or reverse transcription) leading up to integration in the next infected recipient cell. As the Class I/Class II designation of IN mutants is a convenient descriptive apparatus, we first separated the mutants into these broad phenotypic categories based on the straightforward method of direct measurement of the efficiency of mutant reverse transcription-mediated viral cDNA synthesis. Since integration proceeds after the completion of reverse transcription, replication-incompetent mutants that produce high levels of viral cDNA are likely defective specifically for one of the two catalytic processes of integration, and are thus categorized as Class I mutants. The alternative replication-incompetent IN mutant group, those that produce little or no viral cDNA, are blocked for reverse transcription, and *a priori* are Class II mutants.

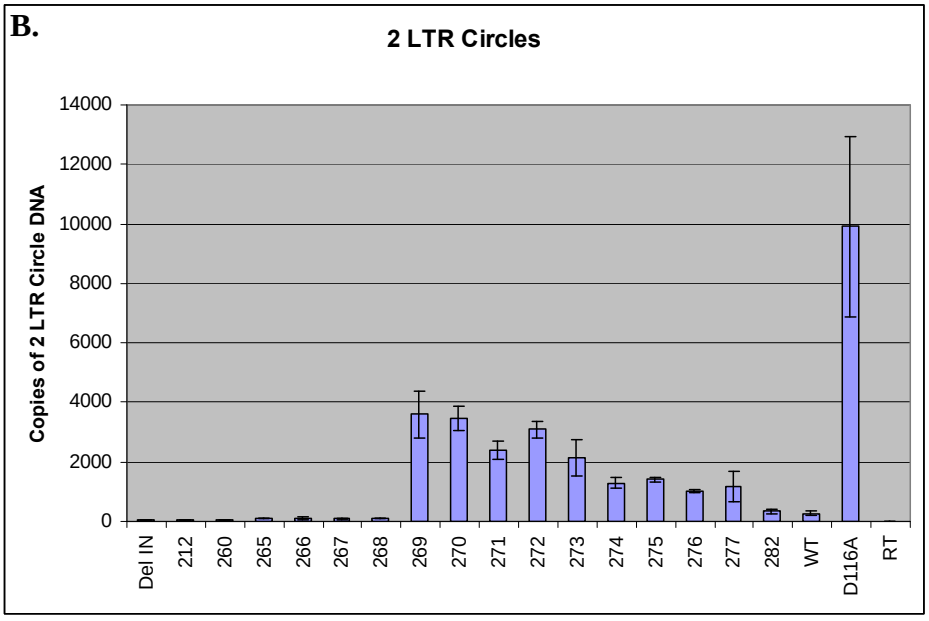
A situation can also arise in which a virus can produce significant viral cDNA, yet is unable to effectively shuttle the PIC components across the nuclear membrane. This is a plausible scenario, since IN has been implicated both in direct and indirect roles for PIC nuclear import. Here it would be misleading to characterize this mutant as Class I, since

the defect is primarily for nuclear import. Thus, to rule out a nuclear import defect as a cause, 2-LTR circles are also quantified. As mentioned, these circular species are used as a surrogate marker for nuclear localization of viral DNA, and its accumulation is symptomatic of Class I mutant infections, as is observed for the prototypical Class I IN mutant, D116A.

The IN truncation panel was assayed for the production of late RT product, which corresponds to the product of an advanced stage of the reverse transcription process, its presence indicative of normal RT function. Cells were infected and total DNA was harvested at eight hours post infection (8 hpi), corresponding to a point of robust viral DNA synthesis in WT infections, as established in our laboratory and in other reports [188]. For the quantification of 2-LTR circles, cells were harvested at 24 hpi, the corresponding point at which the levels of this particular viral DNA species is at its maximum. The results are shown in Figure 18.



8 hpi



24 hpi

Figure 18. Late RT product and 2-LTR circle formation. A) Levels of late RT product, a measure of RT efficiency, were quantified by real-time PCR analysis at 8 hpi. B) 2 LTR circles, surrogates for successful nuclear impart, were quantified at 24 hpi. The IN D116A and WT viruses reveal expected results.

Results-

Guided by the levels of late RT and 2-LTR circle formation in WT and in the prototypic Class I catalytic mutant, IN D116A (Figure 18), the replication and integration defective IN truncations (Figure 14, 16 and 17) can be categorized into Class I or Class II mutants. IN mutants truncated past IN 269 (IN 260, 265-268) can be classified as class II mutants on the basis of their inability to produce significant levels of viral cDNA. In contrast, replication defective IN 270, 271 and 272 viruses produced approximately WT levels of late RT product, and elevated levels of 2-LTR circles. Notably, the circular DNA levels of these viruses did not accumulate to the significant levels produced by the extreme catalytic-defective IN D116A mutant virus. Though these latter viruses (IN 270, 271, 272) did not display as extreme a phenotype as IN D116A, they will be classified as class I mutants, defective primarily for integration.

D116A RECOMBINANTS OF IN TRUNCATIONS:

The discrepancy of integration activities and propensities to produce circular DNA forms in the IN truncations could potentially impact our assessment of the reverse transcription efficiency of the different mutants. This is because the amplicon that is used for late RT PCR analysis is present not only in the late RT product, but also in the linear, circular and proviral DNA forms. A possible confounding factor is that these DNA forms could exhibit unique PCR amplification kinetics thus creating an over-or-under estimation of late RT levels, based on the major DNA form present. Furthermore, integrated DNA is in a protected form, whereas some extrachromosomal DNA species might be subject to degradation (especially linear viral DNA). Furthermore, integrated DNA is duplicated as a component of the host chromosome during interphase S phase replication, and thus proviral products would not be diluted in relation to increasing cellular DNA. In contrast, extrachromosal DNA levels are predicted to be static only for short periods of time, subsequently diluted in relation to cellular DNA with ensuing chromosomal replication and cellular division.

The 8 hour time point for harvesting DNA was chosen so as to lessen from consideration the impact of these alternative DNA fates from the analyses, since circularization and integration peak at later time points. To remove the potential of these complications, we recombined each of the IN truncations with the D116A-inactivating catalytic mutation. Without the possibility of integration, each virus should produce viral cDNA in a quantity predetermined by its inherent efficiency of reverse transcription,

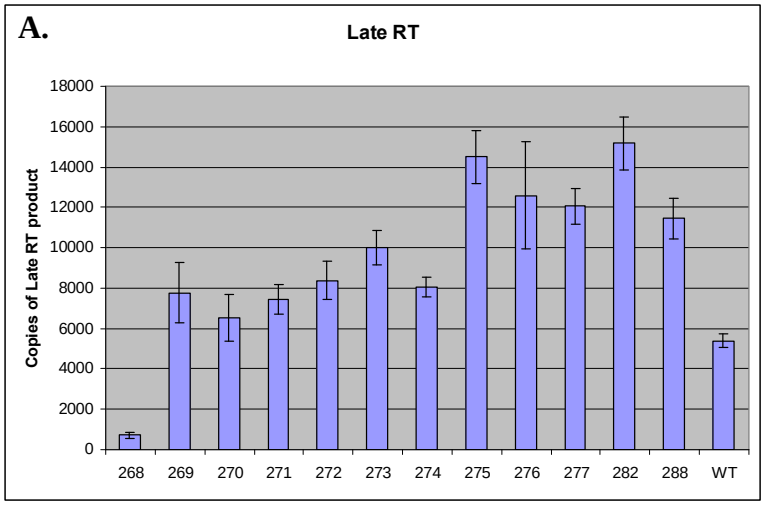
producing extrachromosomal DNAs subject to the same cellular conditions and thus to unbiased scrutiny. The results are presented in Figure 19.

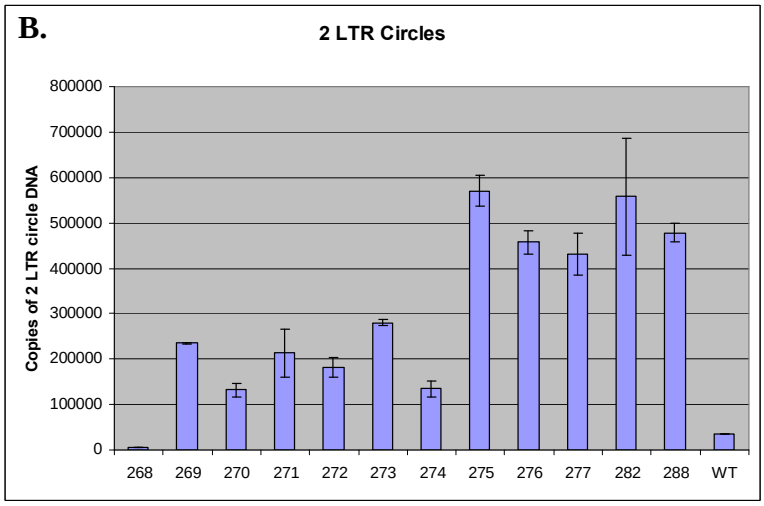
Results-

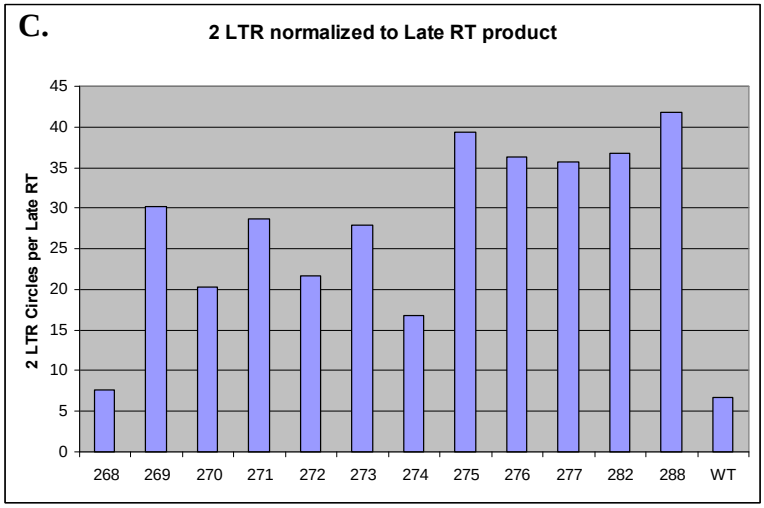
In this experiment, total DNA was harvested at 20 hpi, and analyzed both for levels of late RT product, and 2-LTR circles. The primers used in the late RT PCR will amplify a DNA product synthesized at an advanced stage of reverse transcription, and which is present in all subsequent DNA species, including linear full-length DNA, 1- and 2- LTR circles as well as integrated DNA. Since these viruses are recombined with D116A, integrated forms are removed from consideration, and thus the late RT value will give a measure of total extra-chromosomal DNA levels. The respective profiles of the late RT product formation in Figure 19A corroborate the initial observation that efficient reverse transcription begins with the addition of the natural arginine residue at position 269 (Figure 18A). Full length IN with the D116A mutation (labeled 288 in Figure 19) produces expectedly high quantity of circles. Contrastingly, catalytically-active virus (WT) produces predictably low levels of 2-LTR circles. It is interesting to note two things about these results. Firstly, the levels of 2-LTR circles for the D116A derivatives closely follow the levels of late RT product, indicating that at this time point 2-LTR circles represent a significant fraction of the total DNA in the cell. Since all the viruses are catalytically inactive, blunt-ended DNA products should accumulate to produce 2-LTR circles. The second observation is that the viruses appear to cluster into two groups, based on the levels of 2-LTR circles produced. IN 275, 276, 277 and 282 appear to be

more proficient at making circles, nearly on par with the full-length D116A virus. On the other hand, the shorter molecules, IN 269, 270, 271, 272, 273 and 274, contain fewer circles. This pattern is maintained when the 2-LTR circle levels are adjusted to the late RT values, which gives the relative proportion of 2-LTR circles to total DNA. Though the difference is not striking, it may indicate a slight delay in the ability of IN 269 through 274 to get into the nuclear compartment, perhaps remaining for a longer period of time as linear DNA.

Figure 19. Late RT product and 2-LTR circle formation of recombinant D116A IN truncation mutants. IN truncations were recombined into the IN D116A background to abrogate integration activity, and tested for late RT product formation (A), and 2-LTR circle production (B). The quantity of 2-LTR circles relative to late RT product is displayed in (C).


A.
Late RT
Copies of Late RT product
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
268 269 270 271 272 273 274 275 276 277 282 288 WT


B.
2 LTR Circles
Copies of 2 LTR circle DNA
800000
700000
600000
500000
400000
300000
200000
100000
0
268 269 270 271 272 273 274 275 276 277 282 288 WT


C.
2 LTR normalized to Late RT product
2 LTR Circles per Late RT
45
40
35
30
25
20
15
10
5
0
268 269 270 271 272 273 274 275 276 277 282 288 WT

Summary-

The panel of IN mutant viruses was tested for the ability to efficiently conduct reverse transcription in infected target cells, as a means of delineating Class I IN mutants from Class II mutants. We discovered that robust reverse transcription activity began in IN truncated to residue R269, but was lost for all other truncations N-terminal to this position. Reverse transcription activity was not lost the replication incompetent IN 270, 271 and 272, but rather was as efficient as the replication competent viruses close in length (IN 269 and 273). Thus the replication defective truncations IN 260, 265, 266, 267 and 268 can be classified as Class II mutants that are defective for viral processes other than integration, and IN 270, 271 and 272 can be designated as Class I mutants, blocked specifically at integration (Figure 20)

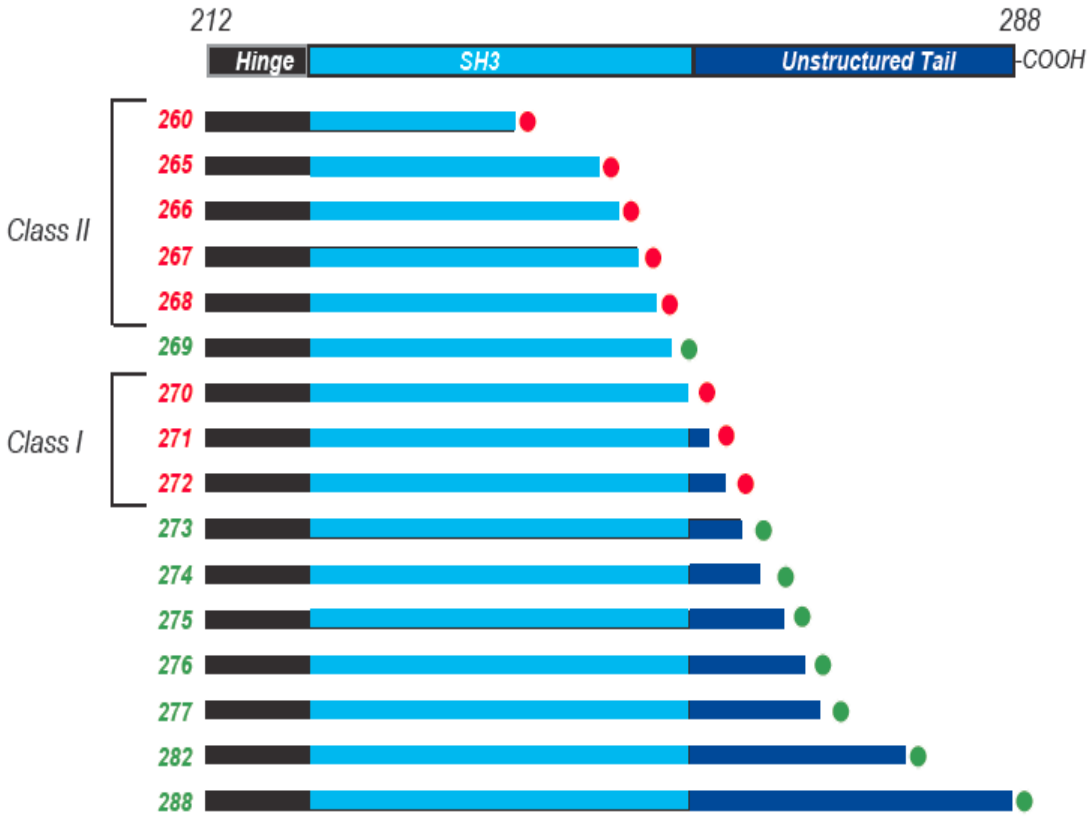


Figure 20. IN truncations fall into two phenotypic groups. A summary of the phenotypic classifications of the IN truncations tested. Green dots indicate replication competent viruses. Red dots denote replication defective viruses. The defective viruses fall into two IN phenotypic groups: IN 260-268 are Class II mutants with pleiotropic effects on viral dynamics, and IN 270-272 are Class I mutants, defective specifically for the integration reaction.

Class I Truncated IN Mutants

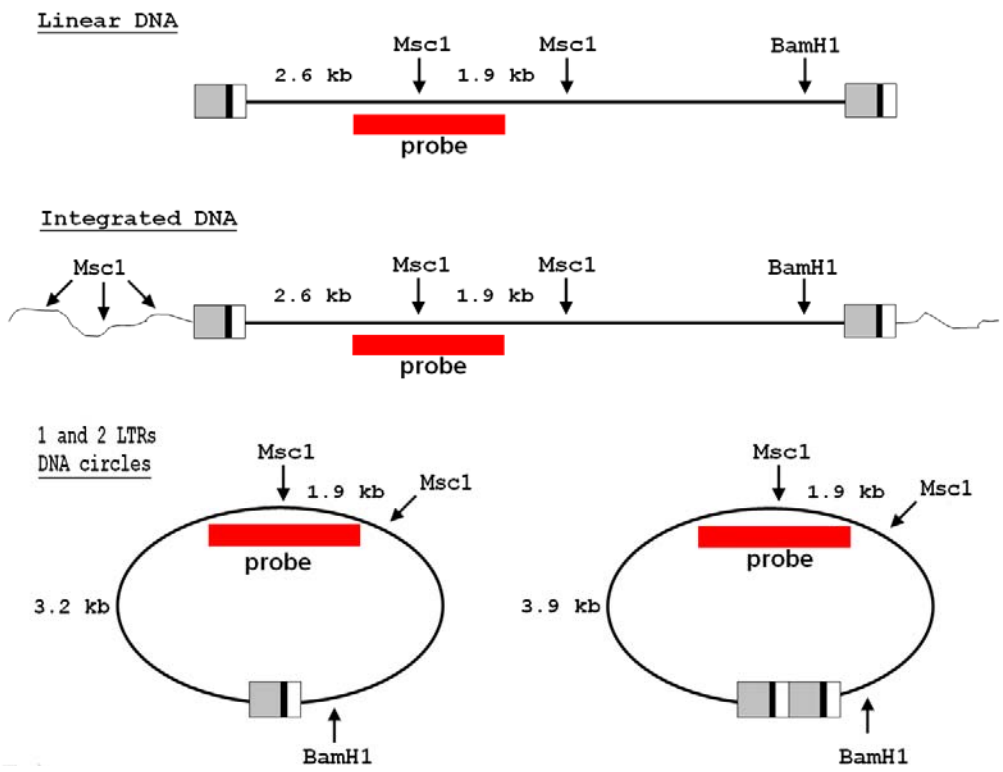
FATE OF VIRAL DNA IN CLASS I MUTANTS:

The observed levels of 2-LTR circle production for the IN truncation panel (Figure 18B) raises the question of the fate of viral DNA in the class I mutants IN 270, 271 and 272: if integration is so severely attenuated, and the quantity of 2-LTR circles does not approach the level observed in IN D116A, then what is the fate of the reverse-transcribed DNA that is being produced at normal levels? Reverse-transcribed viral cDNA can either be integrated, become circularized into extra-chromosomal 1- or 2-LTR species, or persist in its original linear form. Unfortunately, only two (integrated and 2-LTR circles) of the DNA forms can be conveniently gauged by quantitative PCR methods. Integration, as we have shown, can be quantified by genetic means, and by an Alu-PCR real-time-based assay. 2-LTR circles are amenable to specific detection in real-time PCR-based assays by amplifying across the unique amplicon created by LTR-LTR juxtaposition. There is, however, no primer pair combination that can be designed for 1-LTR circles or linear DNA species that would not simultaneously detect all four forms of viral DNA.

To investigate the fate of the viral cDNA as it passes from its synthesis into one of its four possible forms produced during infection, we applied a technique described by Zennou *et al.* [74] (Figure 21A). Here, DNA extracted from infected cells is first digested by specific restriction endonucleases that cleave the viral DNA in a predicted fashion, creating uniquely-sized products derived from the various viral DNA forms. These can

then be resolved by gel electrophoresis, and detected by Southern blot analysis using a labeled oligonucleotide probe that at one end is complementary to a sequence common in all four uniquely-sized fragments. The probe is designed such that it overlaps the cleavage site of one of the endonucleases, and so it can simultaneously detect the level of a second specific viral DNA fragment that is liberated from all the viral DNA forms. This fragment thus represents the total viral DNA presence in the sample. The results are displayed in Figure 21B.

A.



B.

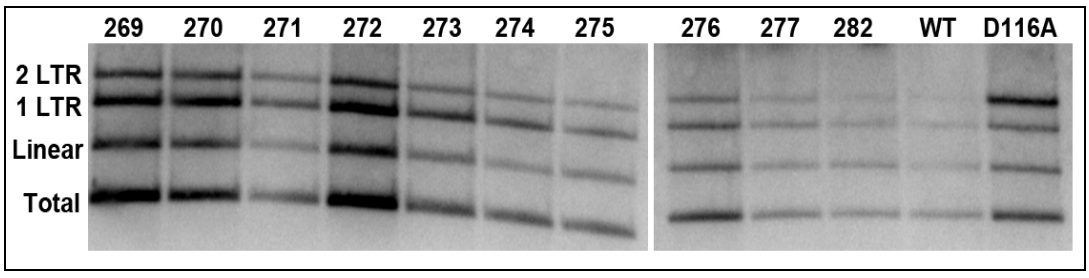


Figure 21. Analysis of the fate of viral DNA in IN truncations. (A) A schematic of the method used to probe the levels of the various forms of viral DNA produced after infection (Adapted from [74]). (B) The levels of viral DNA forms can be observed from the Southern blot analysis of the restriction digested viral DNA present in the IN truncation infections.

Results-

In terms of the relative abundance of each unintegrated DNA product, a trend can be observed proceeding from left to right. Clearly, in the first set of truncations (IN 269 through IN 274), the predominant unintegrated DNA species is the 1-LTR circle. This is especially apparent in IN 269, 270, 271, 272 and 273 and perhaps 274. 2-LTR and linear forms are present in approximately equivalent abundance in these mutants. This pattern changes in mutants IN 275 to IN 277, the abundance of each unintegrated species appear to be approximately equivalent in these viruses. IN 282 and WT display roughly equal levels of linear and 1-LTR DNA but diminished amounts of 2-LTR circles. As expected, the dominant form of unintegrated DNA in the D116A mutant is the 2-LTR circle, with the 1-LTR and linear DNA forms present at a similar level.

These results appear to answer the question of the fate of viral DNA in the Class I IN mutants (IN 270, 271, 272). These Class I mutants do not produce 2-LTR circles with the same efficiency as the typical Class I IN D116A mutant, but instead preferentially produce 1-LTR circles. Since blunt-end viral DNA can be prodigiously ligated end-to-end to produce 2-LTR circles, it is conceivable that these Class I IN truncations are able to 3' process at least one of the viral DNA ends, thus making it a less favorable precursor of the 2-LTR circle, and placing it instead on a preferred path to 1-LTR circle formation. The precise mechanisms by which 1-LTR circles are unknown, and thus it can only be inferred that 3' processed viral DNA is a preferential substrate for 1-LTR circle formation. This conjecture may have some validity, in light of the fact that IN 270, 271 and 272 clearly retain some vestigial integration activity, whereas D116A has none

(Figure 17). The results of the D116A IN truncations also seemed to indicate that IN 269-274 may be delayed in nuclear import of the viral DNA, which could permit more time to process the DNA ends prior to nuclear entry. To follow up on these questions, and others concerning the efficiency of the integration reaction, we were determined to assess the efficiency by which the IN truncation mutants 3' process the viral DNA ends. We thus developed a method of measuring the relative *in vivo* levels of processed viral DNA as described in the following section.

3' PROCESSING EFFICIENCY OF IN TRUNCATION MUTANTS:

Well-established *in vitro* protocols for detecting the 3' processing activity of IN mutants exist, but the artificial conditions of these assays, using isolated IN with viral DNA mimics, often result in discrepant results between *in vitro* and *in vivo* observations. To assay the efficiency of this first step of the integration reaction set directly during infection, we developed a ligation-mediated PCR assay that specifically detects the relative quantities of each of the 3' processed viral DNA ends in infected cells. The protocol is outlined in Figure 22 and in the Methods section, but will be described here briefly: normalized, total DNA extract from cells infected for 8 hours (approaching the peak of late RT product synthesis) is incubated at 16°C in a ligation reaction containing a double-stranded donor DNA substrate, ligase buffer, and *E. coli* ligase. The double-stranded donor DNA substrate is produced by the annealing of a short single-stranded 11 bp oligonucleotide to a significantly longer single-stranded oligonucleotide, a process that is facilitated by the low temperature of the ligation step. The annealed oligonucleotides produce a double-stranded donor substrate that possesses at one end a two bp overhang that is complementary to the exposed dinucleotide on the 3' processed viral DNA end. *E. coli* ligase was selected for its lack of ligation activity for blunt-ended DNA substrates; the enzyme can only ligate together sticky-ended substrates. This characteristic decreases the likelihood of errant and promiscuous ligation of the sticky double-stranded substrate DNA to blunt-ended, unprocessed viral DNA, thus enhancing the specificity of our assay.

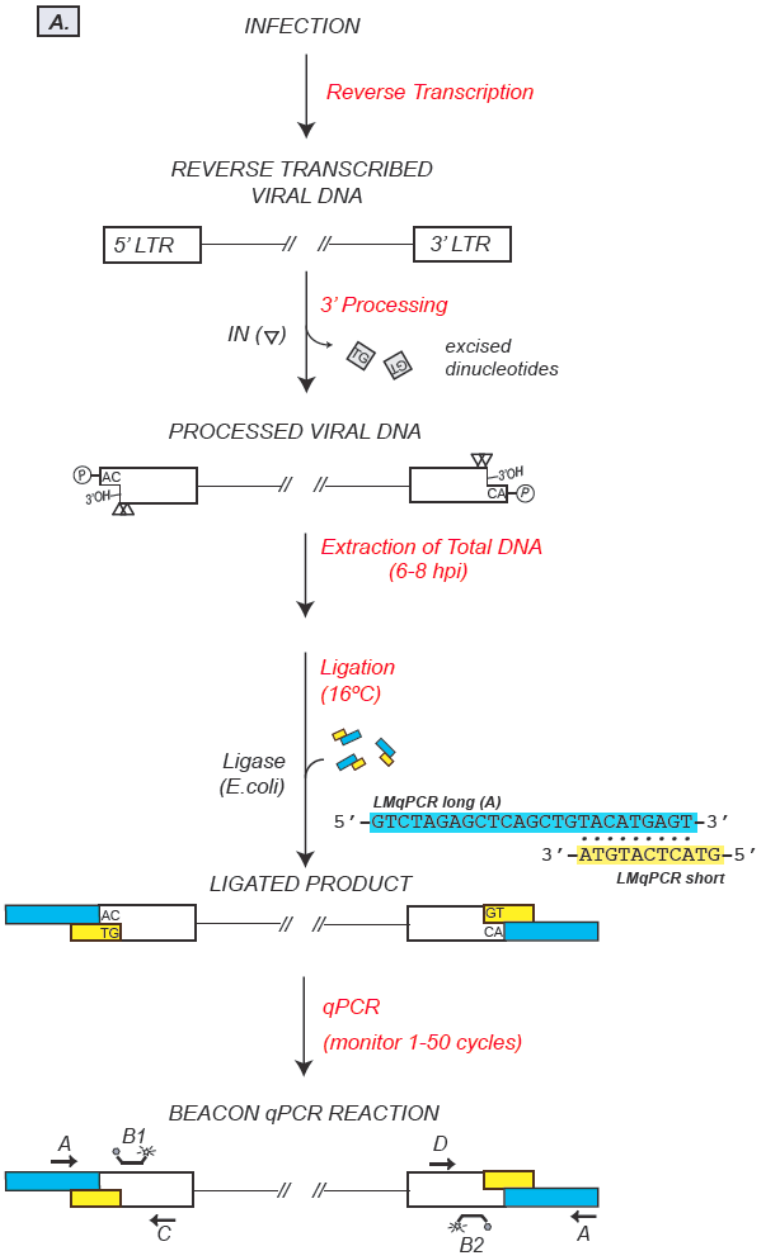


Figure 22. Schematic of quantitative LMPCR 3' processing assay. Details of the protocol are given in Methods.

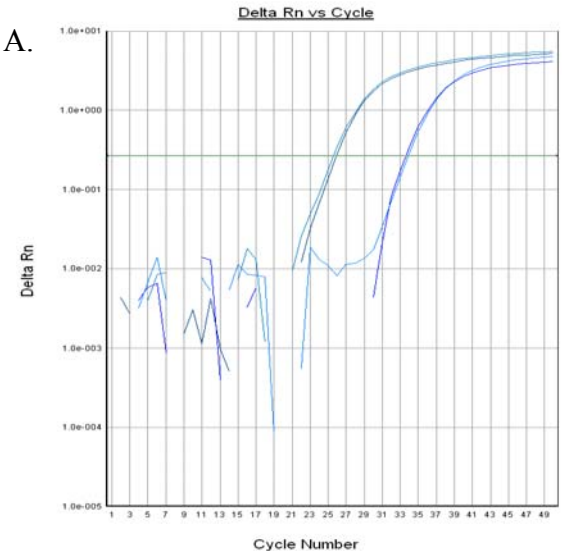
The successfully-ligated products in the ligation reaction are then used as the template in a beacon-mediated quantitative real-time PCR step with primers

complementary to the longer strand of the donor substrate and to an internal LTR sequence. The directionality of the LTRs facilitates the independent investigation of the extent of processing at either end of the viral DNA. The high annealing temperature of the quantitative PCR step discourages annealing by any 11 bp primer carried over from the ligation reaction, which would decrease the efficiency of the reaction.

To satisfy proof-of-principal considerations, the assay was applied in control experiments with the WT and IN D116A virus. WT virus processes the viral DNA expeditiously and provides a strong signal in the assay, whereas IN D116A produces negligible levels of processed product. The extent of end-processing was assayed at both the 5' and 3' LTRs in these pilot experiments, and as the results in Figure 23A and B indicate, the predicted pattern of processing for both viruses was evoked, with IN D116A operating at 0.5% WT efficiency at the 5' LTR, and 2.4% efficiency at the 3' LTR. To ensure the specificity of the ligation reaction, another experiment was conducted in which the sequence of the 2bp overlap of the double-stranded donor DNA substrate was reversed, so as to destroy complementarity with the dinucleotide overlap on the processed viral DNA. The results are shown in Figure 23C and demonstrate that the ligation reaction requires perfect base pair complementation, and thus addresses the assay's specificity. The panel of IN truncations was then examined using this assay, and the efficiency by which viral DNA is processed was determined 8 hpi (Figure 24).

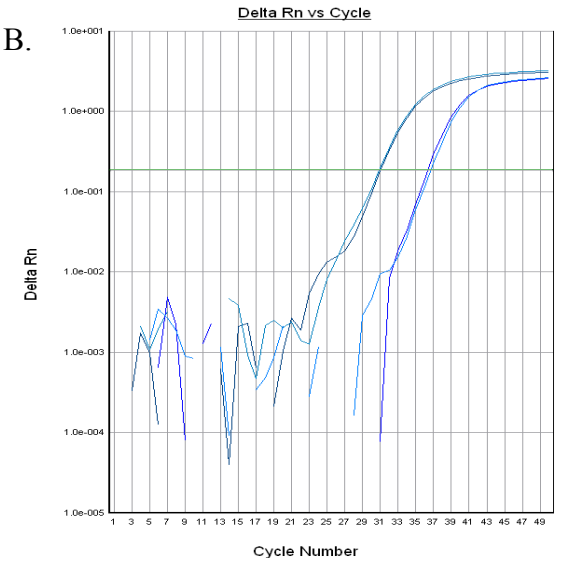
Figure 23. Quality control for LMQPCR reactions. (A) Real-time PCR analysis of the quantity of processed 5' LTR (U3) DNA in WT and IN D116A mutant virus. (B) The same analysis is conducted for the 3' LTR (U5). (C) The specificity of the assay is reflected in the inability of a mismatched dinucleotide overhang on the donor substrate to be ligated to processed viral DNA.

A.



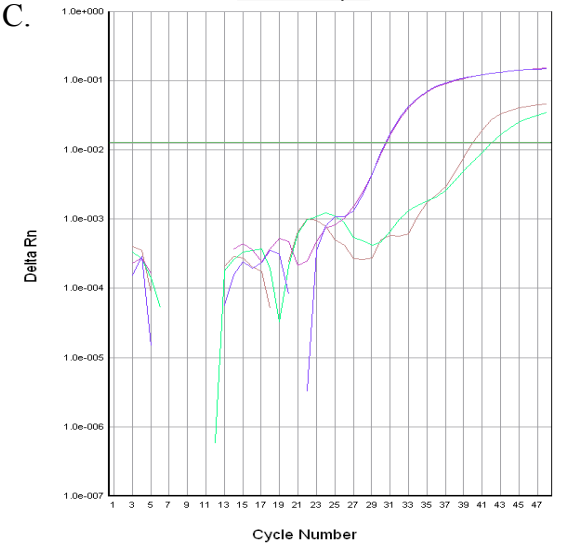
Virus tested	*U3 processing*
WT	***100***
D116A	***0.5***

B.



Virus tested	*U5 processing*
WT	***100***
D116A	***0.5***

C.



Donor Substrate	*U5 processing*
Match	***100***
Mismatch	***0.1***

Results-

Shown in Figure 24 are the levels of processed DNA ends of the IN truncation panel, taken at 8 hpi, and measured at the 5' LTR terminus. Similar results were obtained for the 3' LTR end (data not shown). The pattern of the relative levels of processed viral DNA for the panel of mutants recalls the respective integration frequencies for these mutants (Figure 17). It is evident that a direct correlation can be established between the two. These results indicate that the amount of 3' processing detected for these mutants is directly proportional to the amount of integration as measured by either BSD viral transduction or Alu-PCR assay. Therefore, the lack of integration is best explained by defects at the initial 3' processing step, a step required subsequently for stand transfer and completion of integration. Furthermore, it appears that any molecule that is processed is competent for the strand transfer reaction—3' processing being the limiting step in these mutants. A corollary to this fact is that mutant DNA must be processed at both ends of a given molecule to be a valid substrate for the subsequent strand transfer reaction. Finally, the residual catalytic activity of IN 270, 271 and 272 might be sufficient to avert the over-accumulation of 2-LTR circles observed in the overtly catalytic defective D116A mutant, perhaps with enough activity to skew the fate of reversed transcribed DNA into an increased quantity of 1-LTR circles. Alternatively, since these truncations (IN 270, 271 and 272) possess intact active sites, the IN molecules, though unable to conduct 3' processing, may remain tightly bound to the viral DNA ends, thus occluding them from DNA ligases that would otherwise efficiently ligate the blunt-ended DNA to produce 2-LTR circles.

A.

Mutant	3´ processing (%)	Late RT (%)
268	0	26
269	12	325
270	1	116
271	2	332
272	3	203
273	17	231
274	49	223
275	72	178
276	37	138
277	83	176
282	106	90
WT	100	100
D116	0	248
RT	0	0

B.

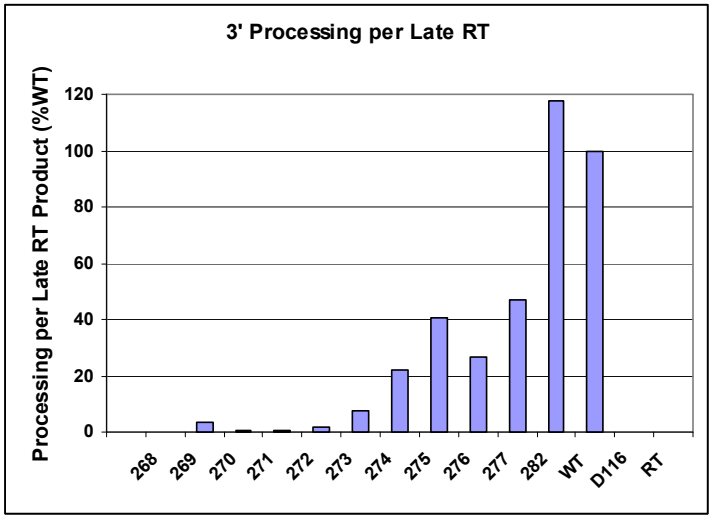


Figure 24. 3' processing of IN truncation panel. Processing of the viral DNA was measured at the 5' LTR in infections of a subset of the IN truncation panel. The results are shown in (A), with the values adjusted to a WT rate set at 100. The DNA samples used for this experiment were also quantified for the levels of late RT product, the results of which are displayed alongside the processing levels. (B) The levels of processing are adjusted for the amount of late RT product in the cell.

Summary/Discussion-

The BSD integration experiment (Figure 17A) indicated a detectable yet almost negligible catalytic activity for mutants IN 270, 271, 272, but from this experiment, which ultimately measures IN-mediated strand transfer activity, it is impossible to gauge the efficiency of 3' processing. Though not commonplace, there is precedence for an IN mutant (IN S119D) that efficiently processes viral DNA, but is drastically deficient for strand transfer activity. In terms of 1- and 2-LTR circle production, the resemblance between the Class I mutants (IN 270, 271 and 272) and the flanking, replication-competent truncations (IN 269 and 273) is undeniable, and thus prompted us to examine in more detail the mechanics of the integration reaction in these viruses. These experiments were driven in part also by the hypothetical notion that processed, unintegrated viral DNA is preferentially manifested as 1-LTR circles, which by Southern blot analysis were observed at similar abundance in the Class I mutant truncation group and the immediately flanking mutants. The subsequent analysis of the 3' processing efficiency did not support this hypothesis since significant processing was not occurring for these mutant. Indeed it was found that Class I truncation mutants, IN 270, 271 and 272 are significantly defective (but not completely) for 3' processing, the first catalytic step of integration. If this is due to an inability to bind to the viral DNA, or if complexes required for catalysis are no longer able to form or other reasons is not currently known, but are subjects for further investigation.

Class II Truncated IN Mutants

We next sought to elaborate our initial 'RT-defective', Class II designation of the IN 260, 265, 266, 267 and 268 (IN 260-268) mutant set to ascertain the nature of the affected viral processes. Given that these mutants were profoundly defective for reverse transcription, we evaluated these mutants for defects in post-integration events. Establishing proviral transcription and protein synthesis as a starting point for the sake of description, the first viral process that is susceptible to IN perturbation is assembly, the process whereby components for the manufacture of virions are amassed at the plasma membrane as a prelude to budding. One symptom of perturbed assembly is a decrease in the efficiency of viral release which can be evaluated by the level of p24 capsid antigen or reverse transcriptase activity in the cell-free supernatant from transfected or infected cells. A gradual diminution of viral release was in fact observed with decreasing IN length (data not shown), indicative of escalating assembly defects. There was, however, little difference between Class II IN mutants and the proximal viable IN 269 virus and the other Class I mutants described above. The next step that can be affected by mutant IN is with regard to the precision of virion particle formation. This defect can be traced to assembly defects, which can misdirect the incorporation of viral components to the virion as it buds. The resulting virions might then display aberrant morphologies, and could be intractable to maturation. Finally, the reverse transcription step is subject to IN-mediated perturbation, either in a direct or indirect fashion. Eliminating any discernable defects at

assembly/maturation can readily identify viruses that are specifically blocked at the step of reverse transcription that are observable during a new round of target cell infection.

INTRAVIRION ANALYSIS:

We conducted the investigation of the nature of the block in the Class II mutants by sequentially backtracking from the reverse-transcription step. We first examined the protein content of the released viral particles, to determine if there was any mispackaging of viral precursors. Poor incorporation of the Pol products, which include RT and IN, could potentially account both for inefficient reverse transcription and subsequent integration. Viral supernatants were filtered, normalized for p24 content, and pelleted through a 25% sucrose cushion. The virus pellets were then lysed and analyzed by western blot analysis using antibodies targeted to specific HIV proteins. The result are shown in Figure 25.

Results-

To substantiate the comparison between levels of intravirion protein content, we ensured that a similar fraction of virions were loaded into each well by probing for the p24 protein, CA, which constitutes the proteinaceous coat of the viral core. The levels of CA are similar in each sample, and little of the p55 Gag precursor or its p41 intermediates were detected, indicating that processing within the virions appeared normal. However, when these blots were subsequently probed for the Gag-Pol products RT and IN, there appeared to be distinct differences between viruses. The levels of IN

and RT incorporation is significantly curtailed in the IN 212, 260, 265, 266 and 267 (IN 212, 260-267). This may account in part for the defect of reverse transcription product

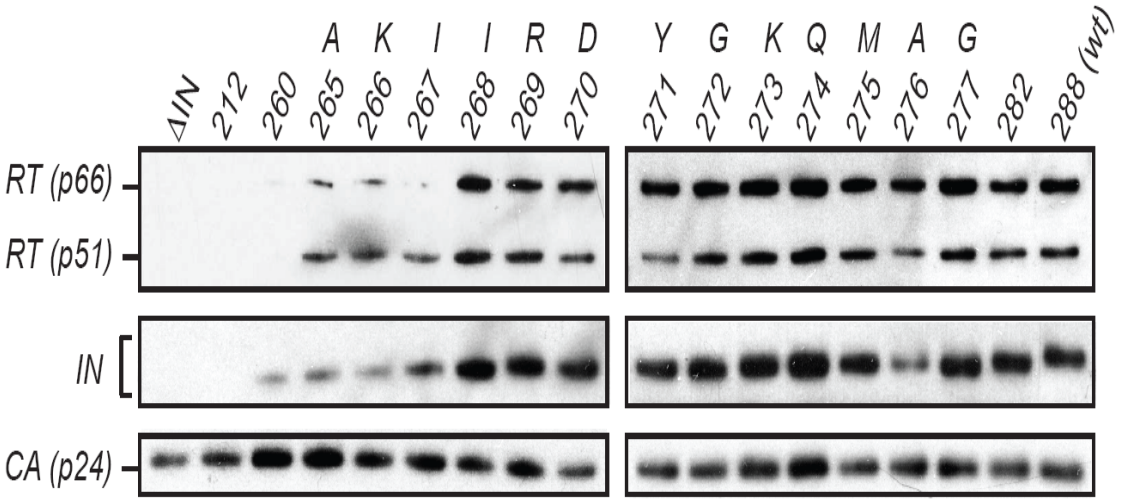


Figure 25. Analysis of virion protein content. Viral supernatants adjusted for p24 content were spun through a sucrose cushion to pellet virions, which were then disrupted and the lysates probed by western blot analysis for intravirion levels of RT, IN and CA. The amino acid sequence of IN is shown at the top.

formation in mutants IN 212, 260-267. It cannot however be discounted that these IN molecules may also be specifically defective for supporting RT. A Pol packaging defect appears to not be the case for IN 268, however, since normal levels of both RT and IN are found in this virus, and this pattern is maintained as expected for the rest of the mutant IN truncation set and the WT virus. Therefore we can tentatively establish that IN 268 is a mutant that is specifically blocked at reverse transcription. However, further analysis of

polyprotein processing and virion maturation of IN 268 is required to give further evidence for this supposition.

INTRACELLULAR POL PROTEIN LEVELS:

As previously demonstrated, mutation within HIV-1 IN can result in premature activation of the viral protease (PR) that ultimately leads to assembly defects [47]. Premature Gag and Gag-Pol precursor polyprotein processing can result in the unnatural accumulation of the mature Gag and Gag-Pol constituent proteins within the cytoplasm. We looked for evidence of this phenomenon in the Class II IN truncation mutants by conducting Western blot analysis of the intracellular protein content of transfected cells. Initially, we investigated the Gag-Pol profile, by probing for two of its processing products, RT and IN.

Results-

The observed intracellular RT levels provide little indication of overt processing defects (Figure 26), as the p66 and p51 RT subunits were detected at approximately similar levels across the viral panel. A slight discrepancy in the RT signal can be detected in the panel to the right, where a side-by-side comparison of ΔIN, IN 212 and IN 260 with the WT is accommodated (these mutants may contain slightly elevated levels of the RT subunits). Other than this slight discrepancy, the RT levels were not overly dissimilar. This, however, is in contrast to the observed IN protein levels, where it was discovered that the Class II mutants, inclusive of IN 268, displayed a significant reduction of intracellular IN. The reason for this was unclear at this moment, but if premature processing was occurring, then the mature IN removed from its protective

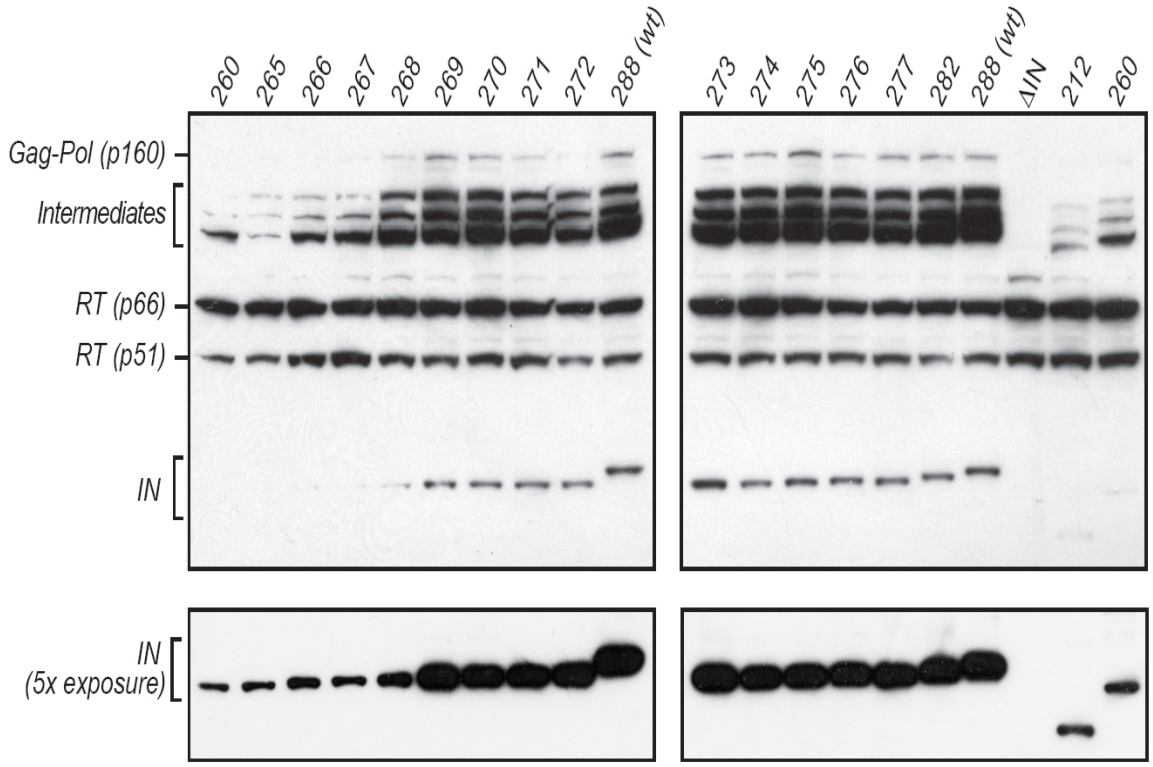


Figure 26. Intracellular levels of Pol products. Lysates of cells transfected for 48 hours with IN truncation plasmids were resolved by SDS-PAGE, and probed for the intracellular levels of RT and IN. A longer exposure is shown in the lower panel to allow detection of the IN protein.

association as a Pol fusion product, would be expected to be degraded in the cytoplasm.

There was a second discrepancy between the individual samples, and this was in the level of some of the intermediate-sized products that ran just below the unprocessed Gag-Pol precursor. Compared to WT, IN 268 is the first truncation to display a WT level of these products. IN 212, and 260-267, contrastingly, displayed minute amounts of these products. It is obvious that it is some form of the Gag-Pol precursor, since they are being detected by antibodies against RT and IN. The next question was whether these

represented degradation products of Gag-Pol, or alternatively, intermediately-processed species produced by cleavage via the viral PR.

INTERMEDIATE GAG-POL PRODUCTS ARE BONAFIDE PROCESSING INTERMEDIATES:

A consequence of probing the intracellular blot with monoclonal α-IN and α-RT antibodies is the detection of the native $Pr160^{Gag\text{-}Pol}$ polyprotein and several intermediate products, as visualized in Western format in Figure 26. The disparity in abundance of these intermediate products among the IN truncations prompted further investigation to determine their origin, and the reasons why they are found in such low quantities in the Class II mutants IN 212, and 260-267. The specific sizes and pattern of the bands in each virus suggested that these were not haphazard degradation products of the Gag-Pol precursor, but rather are specific, partially processed intermediates of the precursor. As such, they would be produced by the specific cleavage action of the viral PR.

We treated transfected cells with a high concentration of the potent and specific HIV-1 protease inhibitor, Ritonavir, and observed the influence of PR inhibition on the levels of the intermediate products. We tested this on four viruses: IN 266 and 267, both representative Class II mutants that contained diminished levels of the intermediate products; IN 269, a replication competent virus that displays a WT pattern of intermediate product formation; and the WT virus, to observe the normal profile in both the presence and absence of the drug. The results are shown in Figure 27.

Results-

In the absence of Ritonavir (Figure 27, lanes 1 through 4), the intermediate products can be clearly detected, and the previously witnessed disparity in their levels between the individual viruses is observed here once more. Similarly, the distinct levels

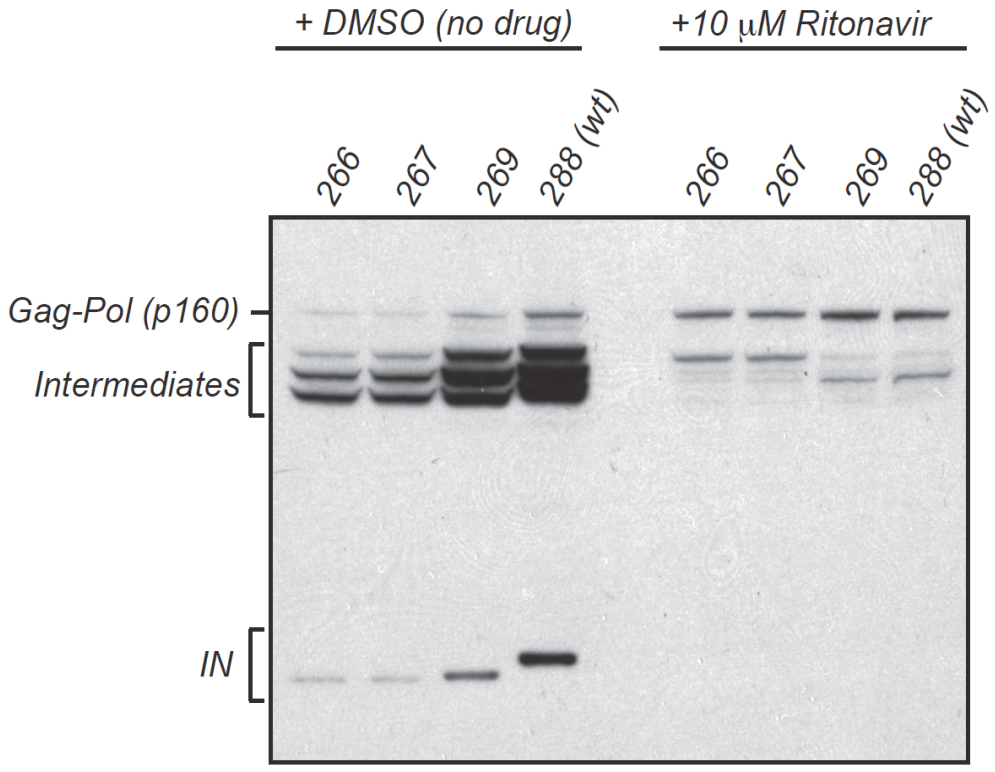


Figure 27. Gag-Pol intermediates are produced by HIV-1 protease activity. The class II mutants IN 266 and 267, and replication competent IN 269 and 288 (WT) plasmids were transfected into 293T cells, in the presence or absence of the PI Ritonavir. Whole cells lysates were collected 48 hours post transfection, and probed for IN in Western blot.

of mature IN are identical to those observed in the previous experiment (Figure 26). With the addition of inhibitor, administered far in excess of the nanomolar range that normally inhibits HIV-1 PR, the intensity of the intermediate products in each virus is significantly

reduced, and the free form of IN becomes undetectable. There is also a clearly perceivable increase in the level of the unprocessed precursor especially in the mutant viruses, which is expected when proteolytic processing is inhibited. It is therefore apparent that the intermediate products are in fact *bona fide*, partially processed fragments of the Gag-Pol precursor.

Close inspection of the small quantity of intermediate products that are produced in the presence of the drug divulges another difference in the Class II mutants. There is the preferential accumulation of a larger than normal molecular weight intermediate (compare Figure 27 lanes 5 and 6 with lanes 7 and 8). Proteolytic processing of the Gag-Pol precursor is not random, but rather is orchestrated through a specified and sequential pattern of cleavages [189, 190]. The addition of the PR inhibitor, as in this experiment, affords the opportunity to observe the initial cleavage reactions of Gag-Pol [49]. The results shown here indicate that in the IN 266 and 267 mutants, the precursors are not only processed prematurely, but also display an aberrant pattern of cleavage.

INTRACELLULAR GAG PROCESSING:

Having identified the intracellular profile of the Gag-Pol precursor, we next investigated the intracellular profile of the Gag polyprotein precursor, probing with an antibody raised against the viral capsid (CA). The range of products this antibody will detect is far less diverse than those for IN and RT antibodies, highlighting only the $Pr55^{Gag}$ precursor, the p41 and p25 intermediate species, and the mature p24 CA. Previous studies conducted in the lab demonstrated that the Gag processing profile pattern of overt Class II mutants and WT viruses tend to converge over time towards the WT phenotype, thus masking mutant defects using the Western blot format. However, harvesting relatively early after transfection (12 hours post transfection) can easily distinguish those viruses that display Gag processing defects. The panel of IN truncations were transfected and harvesting at 12 hours post transfection, the lysis conducted in the presence of Ritonavir to prevent any further possible precursor processing during or after lysis. The results are shown in Figure 28.

Results-

This analysis revealed that though the levels of $Pr55^{Gag}$ and of the p41 intermediate species are similar throughout the panel of IN truncations, the Class II IN truncation mutants (IN 260, 265-268) displayed an irregular p24:p25 ratio, similar to that observed in a virus with gross deletion of the IN region (IN 212, Figure 28, lane 2), itself a well-documented processing defective Class II mutant [47]. The overabundance of the p25 intermediate in relation to p24 has been previously noted as a surrogate marker, a

symptom of premature processing [191], its presence here confirming the premature processing inclination in these viral mutants. The WT virus, as to be expected, displays the most precise control over its processing, exemplified by comparably low levels of p41 and p24. Thus, with the exception of IN 268 in which only one element of processing defects was observed, the Class II IN truncation mutants appear to be impaired primarily at assembly, due to an abnormality of intracellular polyprotein processing induced ostensibly by disregulation of viral protease activation.

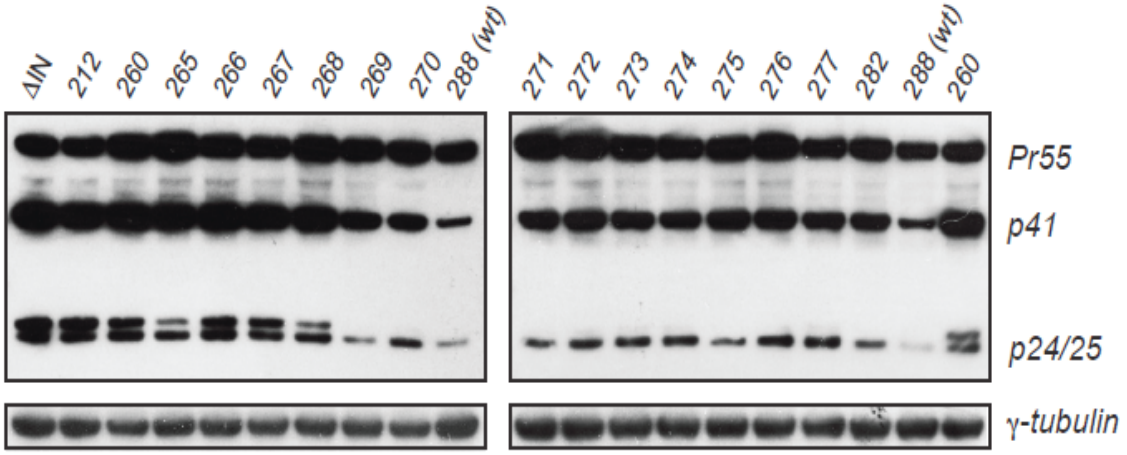


Figure 28. Intracellular Gag processing of IN truncations. Lysates of cells transfected for 12 hours with IN truncation plasmids were resolved by SDS-PAGE, and probed for the intracellular levels of CA, which detects an epitope present in the mature Gag precursor, the p41 and p25 intermediates, and the mature CA (p24). As a loading control, blots were reprobed with γ-tubulin antibody.

Summary/Discussion-

The viral process most vulnerable to loss of control over precursor processing is assembly, since it reliant upon the various biochemical actions of the full-length polyprotein precursors. Assembly defects then directly impact the efficiency of virion output, and the precision of virion composition and architecture. Previous studies have demonstrated that viruses produced with IN deletion or mutation display aberrant intracellular precursor processing profiles, decreased virion release, and produce abnormal viral particles [47, 107]. However, some of these studies usually entail sizable deletions, or almost complete ablation of the IN region. The work presented in this thesis is the first to demonstrate precisely the minimal length of IN at which control of polyprotein processing occurs (IN 269), which is coincidentally the length at which all IN functions are restored, albeit at significantly attenuated levels.

Analysis of the Class II mutants was very informative, especially with regard to the distinct behavior of IN 268. With the original categorization of the IN 260, 265-268, it could not be ascertained at which stage the block existed. As we sequentially backtracked from the common reverse transcription block, initially we discovered that the truncation Class II mutants displayed distinct abilities to package the products of the *pol* gene. The appendage of the isoleucine at position 268 (IN 268) bestowed appropriate packaging of both RT and IN, indicating that the Pol precursor is efficiently packaged into virions. In contrast, IN 212, 260-267 (IN 212-267) contain significantly lower quantities of Pol products on a per virion basis. The reason for the virion protein discrepancy became clear when Western blot analysis of the levels of the intracellular

constituents of Pol were scrutinized. We found a clear disparity in the steady-state levels of the unprocessed Gag-Pol precursor and its partially-processed species. IN 268 displayed a near WT pattern of the partially processed precursors, whereas the other Class II mutants were significantly deficient in these products. Since we were probing against the RT and IN domains which are located at the extreme C-terminus of Gag-Pol, the partially-processed Gag-Pol species represent precursors that have had one or more N-terminal Gag domains removed. It has been reported that a Gag-Pol precursor lacking most of the Gag domain can be efficiently packaged into virions independently of Gag-Gag interactions, but via interactions between the RT domain of the truncated Gag-Pol precursor and the NC domain of the $Pr55^{Gag}$ precursor [192-194]. This phenomenon would then likely account for the ability of IN 268 to incorporate normal levels of Pol products into virions.

The low cellular levels of the Gag-Pol precursor, its intermediate products, and the mature IN in mutants IN 212-267 are indicative either of decreased Gag-Pol synthesis, degradation of the Gag-Pol precursor, or abnormally rapid or aberrant HIV-1 protease cleavage. The first two theories are not likely, since we detected normal to slightly elevated levels of intracellular RT, which is produced by PR-mediated Pol cleavage. The paradoxically low intracellular IN levels can be explained by the fact that cytoplasmic IN, unlike RT, is subject to rapid proteasomal degradation via the N-end rule pathway [195] (for review see [196]). This phenomenon would explain the discrepancy between the levels of cytoplasmic RT and IN. It is possible also, that these IN truncations could be unstable due to improper folding of the CTD, in which case it would also be

targeted for degradation. This question is investigated in the subsequent section. In either scenario, the low levels of IN actually support our hypothesis that the IN 212-267, and IN 268 mutants harbor prematurely processed precursor proteins.

The IN 268 mutant packages normal levels of Pol products into the virion, and contains near normal levels of intracellular partially processed Gag-Pol, but it does display two signs of processing defects. These include the low intracellular IN level, and an aberrant p24/p25 ratio. This mixed phenotypic display indicates that IN 268 may possess a vestige of the necessary function required for controlling precursor processing, an activity that is lost with the removal of additional residues.

Functional Studies of IN Truncations

INHERENT STABILITY OF IN TRUNCATIONS:

The natural amino-terminal residue of HIV-1 IN, produced by viral PR cleavage of the Gag-Pol precursor, is phenylalanine (F). This residue, in conjunction with undefined internal signals within IN, targets the protein for rapid proteasomal degradation by the specialized N-end rule pathway [195]. In the normal course of the viral replication, IN is protected from this pathway by its translation as a covalent adjunct to the Gag-Pol precursor, in which form it is targeted to the sheltered environment of the virion before being released in its mature, processed F-form. It is then afforded further protection during the cytoplasmic transit of the PIC by its intimate associations within the macromolecular complex of the PIC. Mutant viruses that induce premature cleavage of the viral precursors in the cytoplasm of the producer cell would be thought to release free mature IN into the cytoplasm, where it would be subject to expeditious degradation. Such appears to be the case for a subset of the IN truncations, already classified as Class II mutants (IN 212, 260, 265, 266, 267, 268). Indeed, inspection of the intracellular protein profile of these particular truncations reveals significant diminution of cell-associated IN (Figure 26, lanes 1 to 5, compare to lanes 9 to 10). In this case, however, N-end degradation may not be the sole method by which IN can be degraded. As we were manipulating the protein within an ordered region in the SH3 fold, a natural consequence could be that the domain loses the ability to conform to a naturally-ordered structure. Misfolded proteins are also naturally targeted for degradation by the cellular proteasome.

To resolve the question of whether these IN Class II truncations are being targeted to the N-end rule pathway, or are inherently unstable due to improper folding, we investigated the steady-state intracellular levels of N-end rule pathway-resistant IN truncations. As members of our lab have demonstrated, changing the natural N-terminal amino acid to methionine (M) significantly enhances the stability of IN, allowing the protein to be detected far more readily [195]. In this experiment, IN and eGFP are co-expressed from a single bi-cistronic mRNA, a 3'-positioned internal ribosome entry sequence (IRES) providing for the expression of eGFP downstream from the IN sequence. eGFP levels, detected by western blot, serve as the internal control for side-by-side comparisons of IN levels. The results are shown in Figure 29.

Results-

The western blot in Figure 29 shows the steady-state levels of both eGFP, serving dual roles of a loading and expression control, and a small panel of our IN truncations. A subset of these was observed to have significantly diminished cytoplasmic levels in producer cells. To distinguish between their degradation by the specialized N-end rule or the misfolded protein pathways, we made N-end rule resistant recombinant IN molecules by appending a protective Met residue (M-form) to the N-terminus. As the Western blot clearly demonstrates, the steady-state levels of each IN truncation are nearly equivalent for IN 265, 267 and 268, with only IN 260 and 266 displaying decreased IN levels, indicating that these two truncations maybe unstable. Though this is only a snapshot of the individual protein levels, and does not divulge of the respective half-lives of the

truncations, there appear to be little difference in the steady-state levels of the IN 267 and 268 truncations compared to WT IN. This indicates that blatant misfolding of the CTD may not explain the observed low cytoplasmic levels of IN in virus producer cells. (Figure 26).

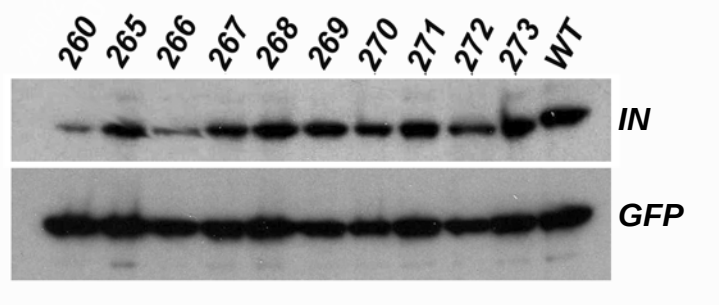


Figure 29. Stability of M-IN. Methionine-initiated IN truncations were tested for cytoplasmic stability. eGFP and IN are translated from a bi-cistronic transcript, and so the levels of GFP serve as an internal control for the levels of IN.

D116A COMPLEMENTATION OF IN TRUNCATIONS:

A possible explanation for the observed phenotypes of the progressively deleted IN CTD mutants is simply that the enzyme becomes increasingly unstructured and misfolded as it is gradually shortened. This is especially pertinent when removing residues past IN 270, the outer margin of the single structural element in the CTD, the SH3-like fold (Figure 7C). Indeed, in our panel of mutants, IN 268 and shorter are the only mutants that exhibit multiple defects in viral dynamics. If in fact the pleiotropic effects observed after truncating past 269 are caused by the loss of one or more specific CTD-contributing activities, and not by gross protein misfolding, then those activities could potentially be supplied *in trans* by another IN protein that has an intact CTD. We thus tested the IN truncations in a reciprocal rescue scheme with the IN D116A mutant, which contains a mutation in the CCD, but has an intact CTD. Interdomain complementation is a commonly used methodology to discriminate pair-wise patterns of IN dependency, due to the integration reaction requiring distinct roles from participating monomers. In one model based on recent CCD-CTD crystallographic data, the viral DNA processed end is in complex with the CCD of one monomer while the same viral DNA is coordinated by the CTD of the opposing IN monomer (Figure 10). The IN truncation mutants possessing a properly-folded and thus catalytically active CCD, should complement the D116A IN, and in turn the D116A IN with its intact CTD should complement the replication defective CTD truncation mutants. Hybrid IN viruses made by co-transfection of mutant plasmids were tested for integration frequency and *in vitro* 3' processing activity. The results are shown in Figure 30.

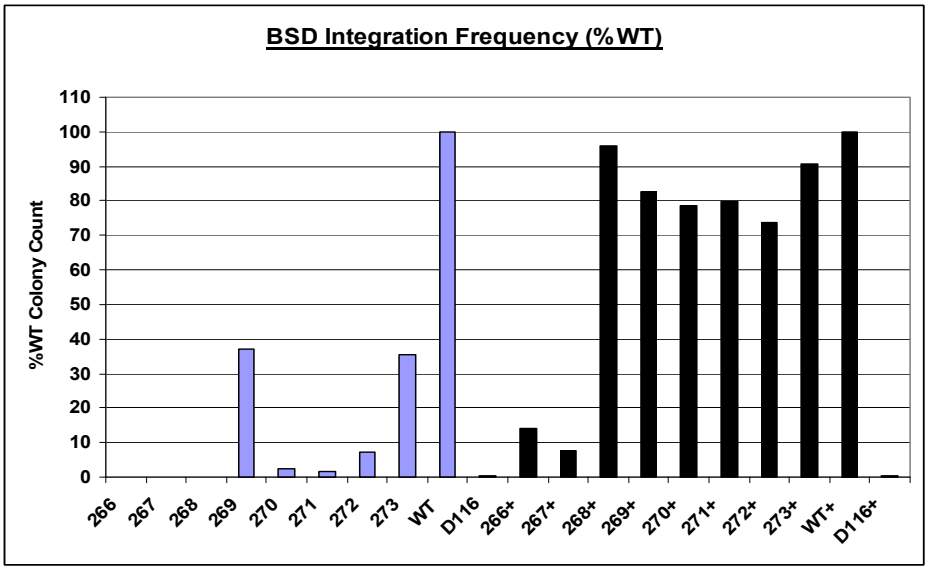
Results-

The integration frequencies of the single and hybrid viruses are shown in Figure 30A. The single viruses each displayed a rate of integration per p24 unit similar to that previously observed. For example, IN 266 and 267 have no detectable transduction ability, while IN 269 and IN 273 are capable of more than 10% of the WT rate with the intervening IN 270-272 viruses being severely attenuated. When co-transfected with an IN D116A viral plasmid to make hybrid IN viruses, a clear demarcation between viruses that could or could not be efficiently rescued for integration was established. IN 266 and 267 were rescued by IN D116A, but only minimally, achieving 14 and 7% WT transduction rates respectively. Contrastingly, IN 268 was rescued by IN D116A very efficiently, approaching the WT level of transduction efficiency. The Class I mutants IN 270-272 were also efficiently rescued by IN D116A, likewise attaining WT transduction efficiency.

The efficiency of processing at the one end of the viral DNA (5' LTR) was then assessed; the results are shown in Figure 30B. In the single viruses, the efficiencies of processing are as they were previously determined (compare with Figure 24). Based on the complementation data for the BSD integration assay, mutants IN 266 and 267 were expectedly and significantly rescued by D116A for processing, but only minimally. IN 268 however displayed in this assay, as in the BSD assay, abundant integration activity in the presence of IN D116A. Notably, the 3' processing reaction in the IN 269 and IN 273 hybrid viruses was significantly enhanced by the IN D116A mutant, achieving an

approximately 5-fold increase compared to the 3' processing rate encountered in the uncomplemented viruses.

A.



B.

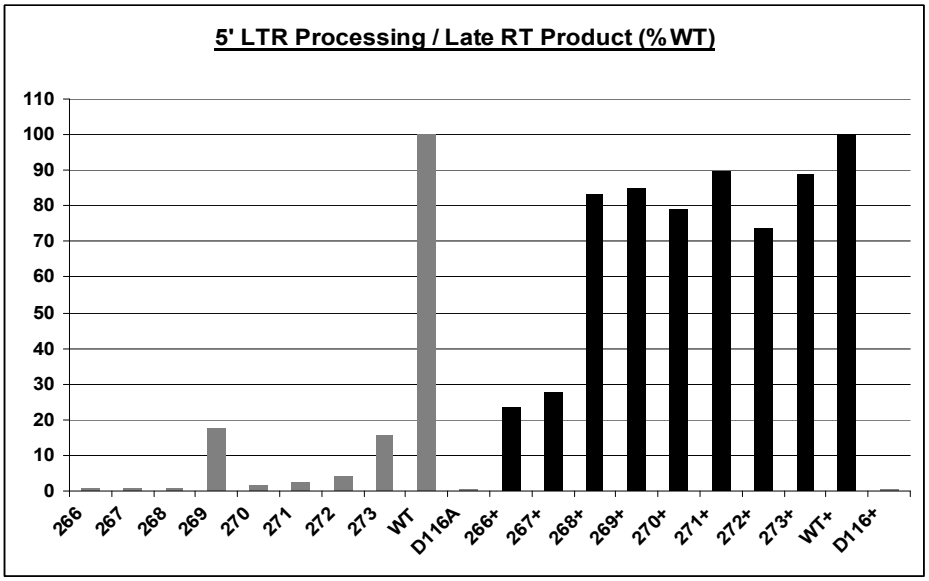


Figure 30. Complementation of IN truncations with the CCD Class I IN D116A mutant. One-step single or hybrid viruses made by cotransfection of the virus expression plasmids for the IN truncations or IN D116A were tested for integration reaction activity in the BSD integration assay (A), and for efficiency of 3' P (B). + = Hybrid D116A virus.

ATTEMPTED RESCUE OF IN 267 AND IN 268 WITH CLASS I CTD TRUNCATION MUTANT:

It was established in the previous experiments that the canonical Class I mutant, IN D116A, effectively rescues IN CTD truncations from both phenotypic classes: IN 268 (Class II), 270, 271 and 272 (Class I). While these Class I mutants are clearly rescued for a defective integration activity, the basis for IN 268 rescue is ambiguous. The IN 268 mutant cannot facilitate efficient reverse transcription, but we cannot extrapolate from this that it is also defective for integration activity. IN D116A rescues the reverse transcription defect of IN 268 (data not shown), but is it necessarily assisting IN 268 also for integration? It is possible that given a cDNA substrate, IN 268 can proceed efficiently with integration without any form of complementation.

Class I IN mutants located within the CTD are uncommon, but our studies have led to the identification of three examples, all consecutive truncation mutants: IN 270, 271 and 272. These mutants offer a unique opportunity to study the rescue of the reverse transcription defect of IN 268 but without their contribution to- or complementation of- integration. If the isolated IN 268 is integration defective, its mutation, being harbored in the CTD, will require a distinct IN bearing an intact CTD for rescue. To investigate this possibility IN 267 or 268 viral plasmids were co-transfected with the Class I IN 270 mutant viral plasmid to make hybrid viruses, which were tested for successful complementation by measuring transduction efficiency in the BSD assay. The results are shown in Figure 31.

Results-

It was previously shown that packaging a catalytically-inactive IN into virions is sufficient to rescue the RT defect of IN Class II mutants [172], a finding that we have corroborated (data not shown). We co-transfected IN 270 with IN 267 or 268 to determine if the contribution of reverse transcriptase activity was sufficient to rescue these Class II mutants. The results (Figure 31) clearly indicate that neither IN 267 nor 268 can be rescued simply by co-transfecting an IN that supports reverse transcription. Thus it appears that the nature of the disruption of the CTD in IN 268 affects not only the majority of secondary IN functions, but also of the primary integration functions.

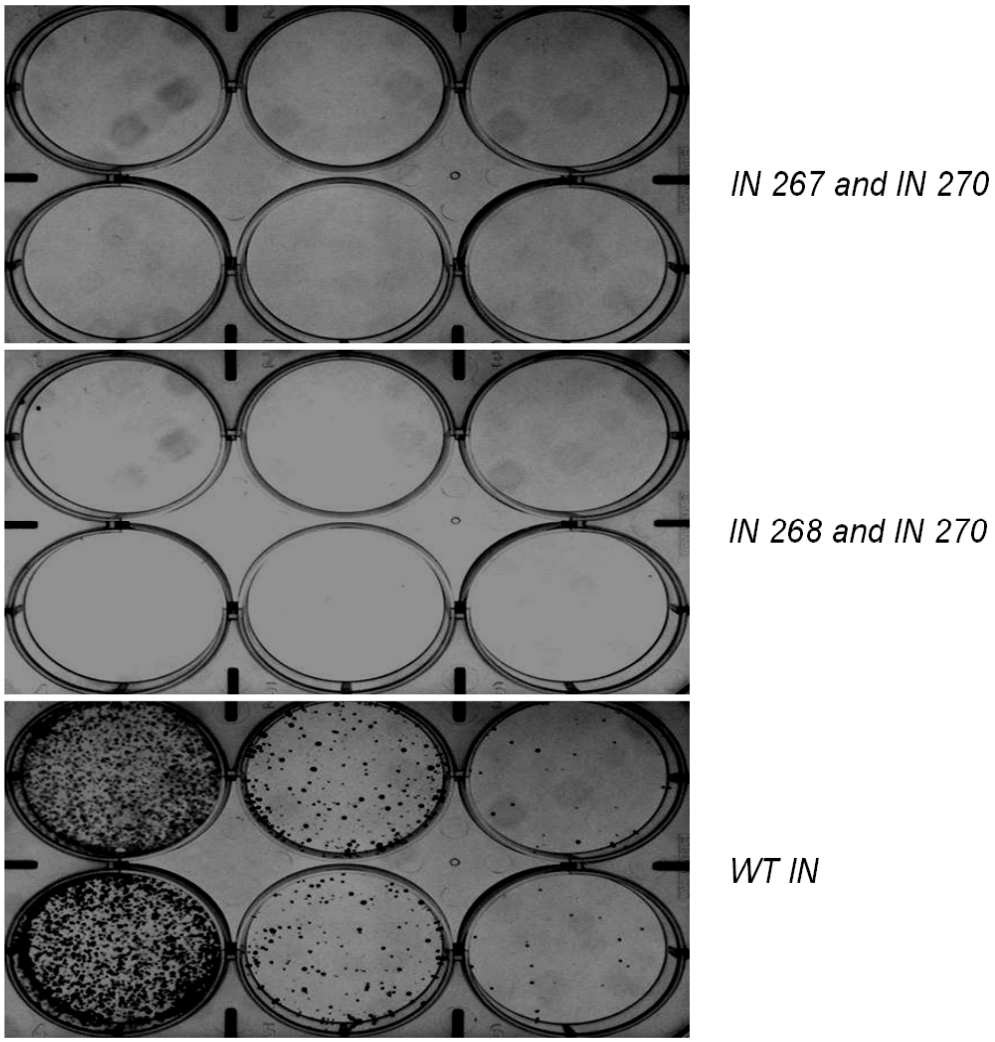


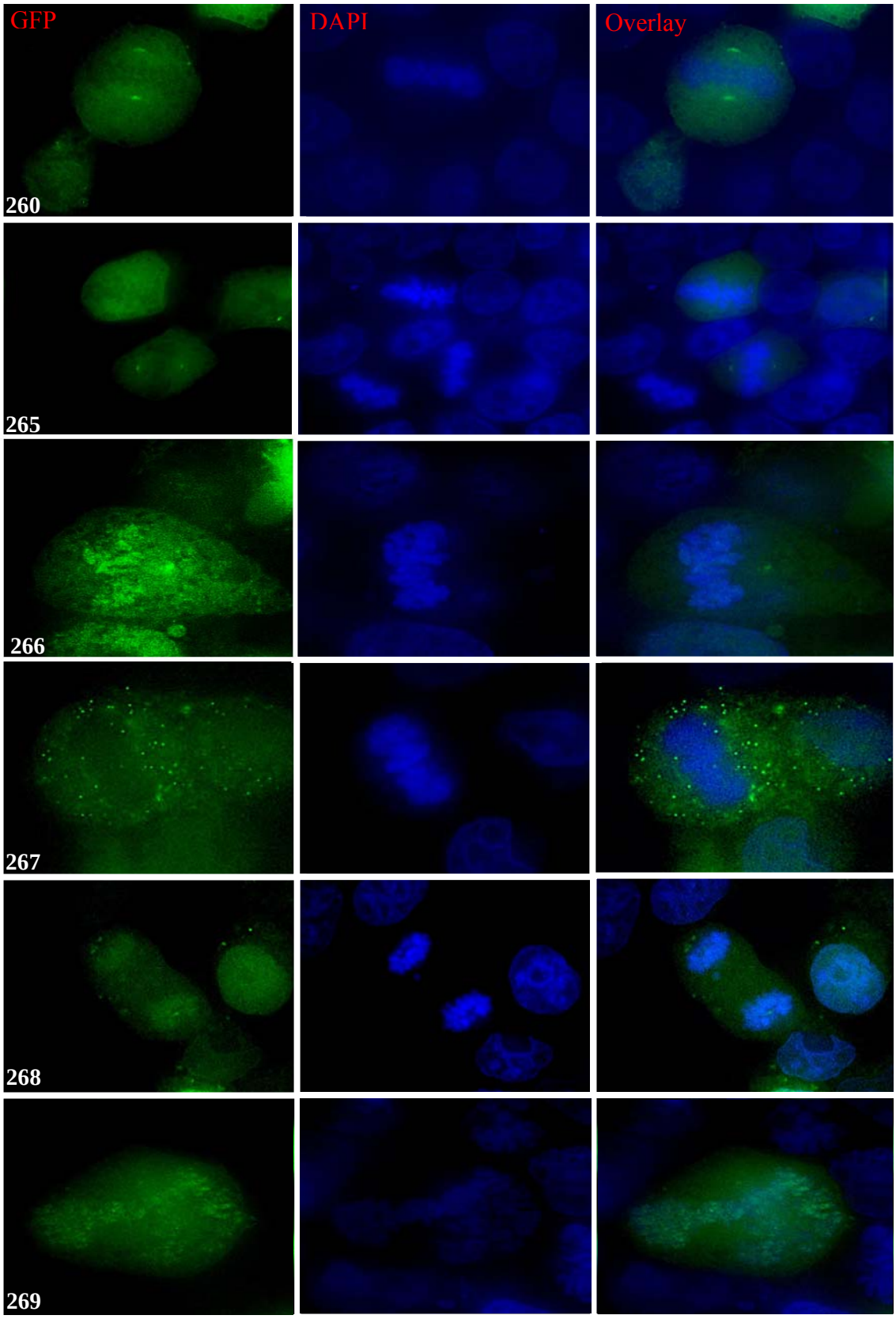
Figure 31. Complementation between Class I and Class II IN truncations. Hybrid IN 267/270 and IN 268/270 viruses were tested for successful complementation in the BSD integration assay. Full-length IN virus (WT) was used as a positive control. Inoculated cells were split into selection in consecutive wells at decreasing densities.

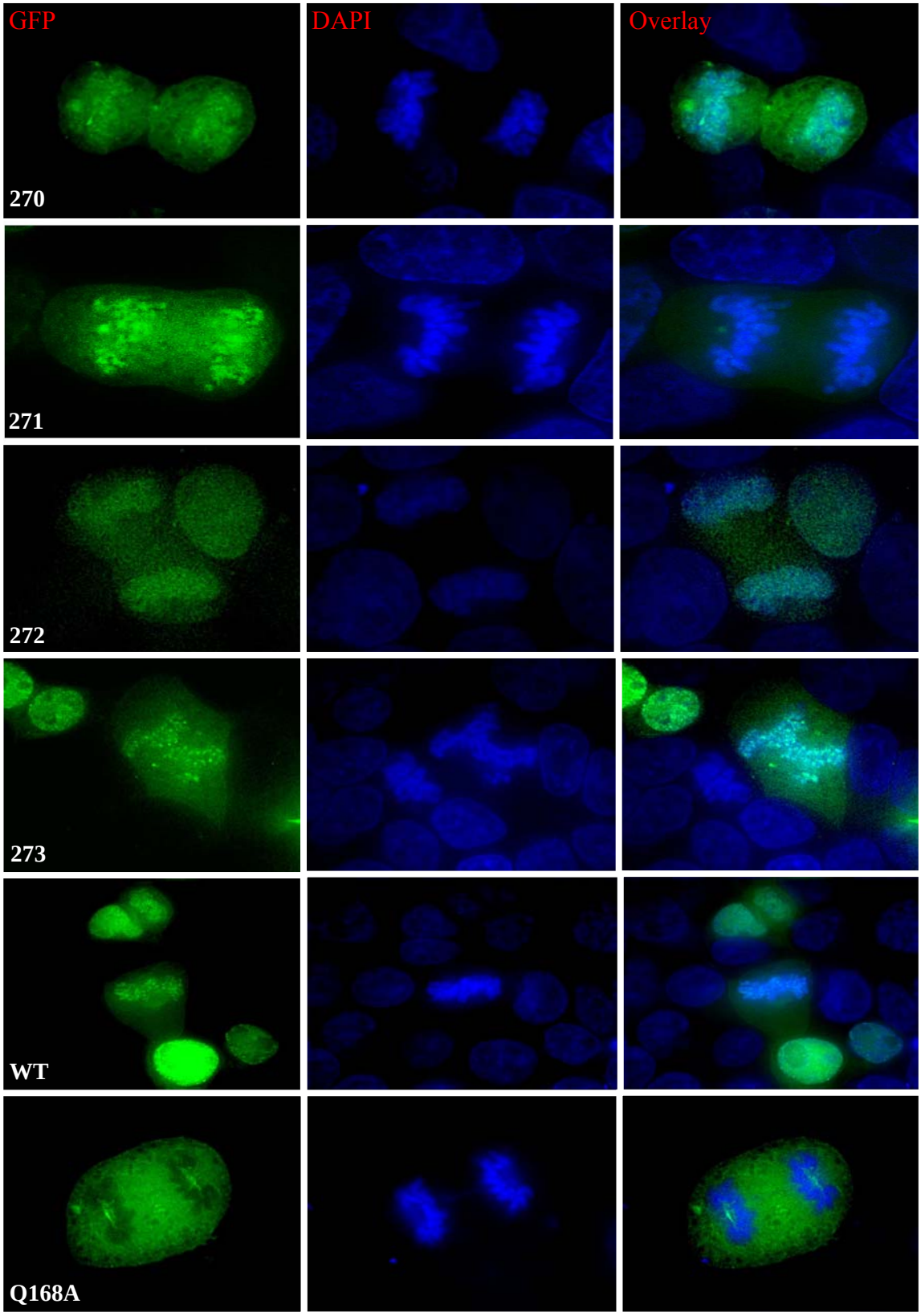
INTERACTION OF IN WITH LEDGF:

LEDGF/p75 was first identified as a cellular binding partner of IN by co-immunoprecipitation [167], and the interaction was subsequently confirmed in a yeast-two hybrid assay ([140]. Disrupting the interaction by siRNA knock-down of endogenous cellular LEDGF or by mutating key residues within the LEDGF binding site in IN results in several-fold defect in transduction ability, the block identified as occurring at integration [197]. Studies indicate that the IN-LEDGF interaction plays several potential roles including nuclear import of the PIC, tethering IN to the chromatin and enhancement of IN catalytic activity [140, 197, 198]. The interaction is mediated via residues within the IN catalytic core, and a single amino acid substitution (Q168A) in this domain can abrogate the interaction. The role of LEDGF as a chromatin tether for IN is particularly apparent during mitosis, wherein IN is found intimately associated with the condensed chromosomes. This chromatin staining by IN is lost in cells that lack LEDGF, or with IN mutants harboring the Q168A mutation that abrogates LEDGF interaction.

Since the IN-LEDGF interaction is a well-characterized and essential phenomenon, we tested the ability of a subset of our IN truncations, in the context of C-terminal fusions of eGFP, to stain condensed mitotic chromatin, using this as a visual marker for LEDGF interaction. It was hoped that this line of inquiry would give an inclination as to the overall structural integrity of the truncated enzymes, that an IN-LEDGF interaction mediated by IN CCD residues should not be disturbed by CTD mutations, unless those mutations exert global effects on the mutant protein. The results are shown in Figure 32.

Figure 32. IN truncation and LEDGF interaction. eGFP-IN truncation fusions were transfected into 293T cells, fixed 12 hours post transfection, mounted onto slides, and observed for chromatin binding, a surrogate marker for LEDGF interaction.


GFP
DAPI
Overlay
260
265
266
267
268
269


GFP
DAPI
Overlay
270
271
272
273
WT
Q168A

Results-

The association of HIV-1 IN with mitotic chromatin is a surrogate marker for binding to LEDGF, which acts as a molecular tether between chromatic DNA and IN. A subset of IN truncations was fused at the N-terminus to eGFP and then assayed for the ability of the eGFP-IN fusion to adhere to mitotic chromatin. The LEDGF interaction with IN is mediated by residues in the catalytic core of IN, and it is thus unlikely that this association would suffer from a CTD mutation. LEDGF-IN association if disrupted, however, may be indicative of global protein effects. This line of inquiry is part of our effort to ascertain the specific defects of these IN truncations.

The results in Figure 32 demonstrate that some of the IN truncations (IN 260, 265, and 267) displayed a chromatin staining pattern similar to the LEDGF-binding mutant, IN Q168A. In the GFP channel, a shadow can be discerned where the condensed chromatin is positioned. This so-called negative staining is indicative of a lack of chromatin association, and by proxy, a lack of LEDGF binding. The first truncation to display consistent chromatin binding was IN 266, suggesting that this mutant can functionally associate with LEDGF. The next truncation, IN 267, however, formed large cytoplasmic aggregates that demonstrated no affinity for condensed chromatin. The next truncation IN 268, and the remainder, IN 269, 270, 271, 272 and 273, were all found to bind chromatin, though the level of binding intensity displayed by full length IN was only consistently observed in IN 273. In some cases, we could however find a few examples in which IN 269 or 270 fusions did not appear to be associating with the mitotic chromatin (data not shown). This discrepancy at this time cannot be accounted for. It would be

instructive in future experiments to test each truncation in the background of the Q168A mutation, which may serve to dispel these ambiguities.

In summary, we discovered that eGFP-IN 266 and 268-273 bound to condensed chromatin during mitosis, indicative of LEDGF-binding, whereas IN 260 and 265 did not. IN 267 displayed a more drastic defect, producing cytoplasmic aggregates that may deter DNA binding. These results provide more support for IN 268 being the minimal length of IN after which secondary IN functions are constrained.

Selection of Tail-less Second Site Revertant Viruses

HIV-1 is a highly adaptable virus, evolving rapidly in the face of selective pressure to constantly achieve the fittest form within the context of a given biological environment. It is facilitated in this regard by the large number of progeny, its rapid replication rate, and perhaps most significantly by the error-prone nature of its genomic replication. These features form the basis for the ease with which the virus constantly evades the dynamic adaptive immune responses of the host, and for the genesis of drug resistant viruses in patients. In tissue culture, however, this aspect of HIV-1 has been indispensible for studying the mechanisms of antiviral drug mechanism and inhibition, and for eliciting the significance of viral sequence and structural elements. Selective pressure can be applied externally in the form of suboptimal concentrations of antiviral drugs, which permits a low level of viral replication and thus the accrual and subsequent selection of fitness-enhancing mutations that confer drug resistance. Selective pressure can also be applied internally by introducing mutations into areas of the genome that attenuate viral replication. Often the virus simply reverts to the wild-type sequence, but sometimes second site mutations can arise which can restore viral replication. Such studies have been successfully applied to deciphering the relevance of RNA secondary structures in the TAR element, and for identifying hidden intraprotein interactions.

As a means of understanding the function of the IN tail, we used this technique of forced evolution to elicit second site revertants in a tail-less IN mutant. We used the IN 269 stop mutant in these selection experiments, since it is completely devoid of the

unstructured tail, and as we have demonstrated is capable of sustained, albeit significantly attenuated replication. The sluggish replication kinetics of the founder virus supplies the impetus for the selective advantage and outgrowth of any fit variants that may arise. To encourage the genesis of second site revertants, we ensured that reversion to the wild-type sequence would be unlikely, by introducing five stop codons after the 269 residue: three in the IN reading frame, and one in each of the two overlapping reading frames (Figure 33). The CEM cell line was selected as the tissue culture system through which the virus would be passaged for several reasons. Firstly, these cells are grown in suspension, and as such are amenable to maintenance in large cultures that can achieve high cell densities. This effectively provides large numbers of cellular targets for multiple rounds of infection, thus favoring the accrual of random mutations. Secondly, these cells are permissive, not requiring Vif function for productive infection, and thus eliminates any possible concern engendered by the unintentional disruption of the *vif* ORF created by the introduction of the five stop codons. Finally, the outgrowth of revertant viruses can be readily perceived by conventional microscopic detection of a sudden uncharacteristic burst of robust syncytia formation.

The five-stop codon IN 269 sequence [IN 269 (5x)] (Figure 33) was engineered into two proviral constructs. The first construct encoded functional viral proteins except for Vpu (Vpu-). A *vpu*- background was warranted since *vpu*- viruses proliferate more efficiently and create syncytia more readily in CEM cells than their *vpu*+ counterpart (data not shown). This may be due to an enhancement of cell-cell transmission by the accumulation of virus particles on the cell surface. The second proviral construct has a

frame-shift mutation within the *vpr* gene such that the Vpr protein is unstable and effectively absent intracellularly; all other genes are intact in this construct including *vpu*. Vpr was disrupted in this virus to prevent cell cycle arrest of infected cells and to allow viral proliferation and spread. This virus is predicted to spread preferentially by the free virus route, rather than through cell-cell contact. CEM cells were infected with a low inoculum and subsequently passaged at regular intervals. At every fourth or fifth split, the infected cells were massed by centrifugation and the cell-free supernatant containing cell-free virus transferred to fresh, uninfected CEM cells. Occasional cell-free transfer of progeny virus to uninfected cells was done so as to increase the likelihood of selecting revertant viruses while culling out unfit viral species (i.e., cell-cell transmission might favor the persistence of the attenuated founder virus). The scheme is shown in Figure 33.

Robust syncytia formation was detected approximately four months after the initial infection with the first virus (*vpu*-) and approximately one month later for the second viral construct (*vpr*-). Total DNA from a portion of the culture was then harvested at weekly intervals and regions of interest were amplified by PCR, and the DNA examined by bulk PCR sequencing. Although chromatographic DNA sequencing of an aggregate PCR product from a diverse viral population under growth selection will only detect major alterations (reliably >20% deviation), we were fortunate to uncover a predominant codon altering substitution mutation as well as less representative but linked substitutions that changed the IN amino acid sequence (Figure 34). To test for the validity of the revertant phenotype, mutagenic primers were designed to assemble test proviral

constructs that were isogenic for all viral sequences except at the relevant, revertant locus.

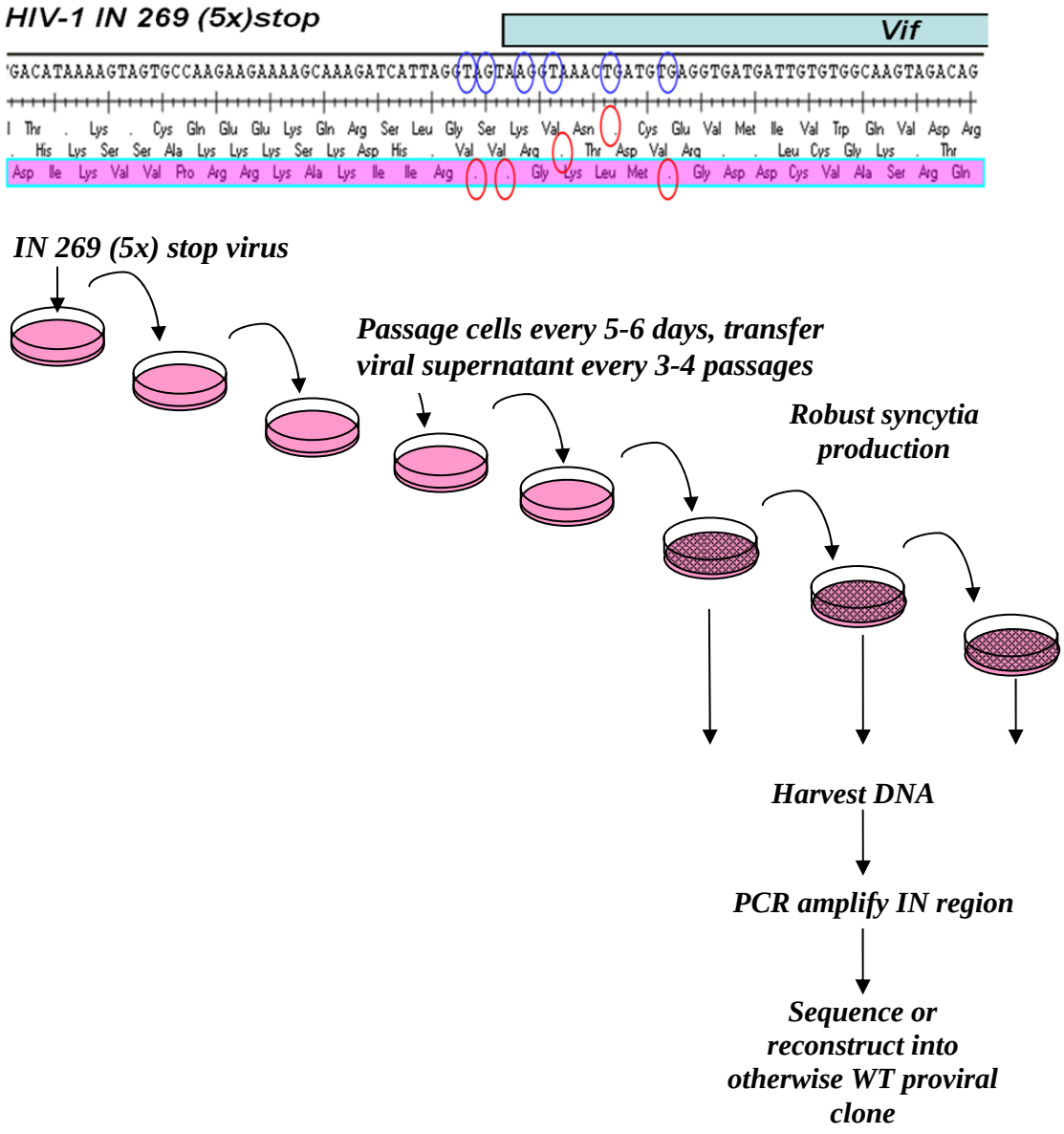


Figure 33. Selection of viral revertants. Schematic of the procedure used to procure viral revertants of the IN 269 stop virus. To prevent reversion to WT length, five stop codons were introduced after the 269 codon, two in alternative reading frames, and two extra ones just after the first, with one eliminating the 273 residue.

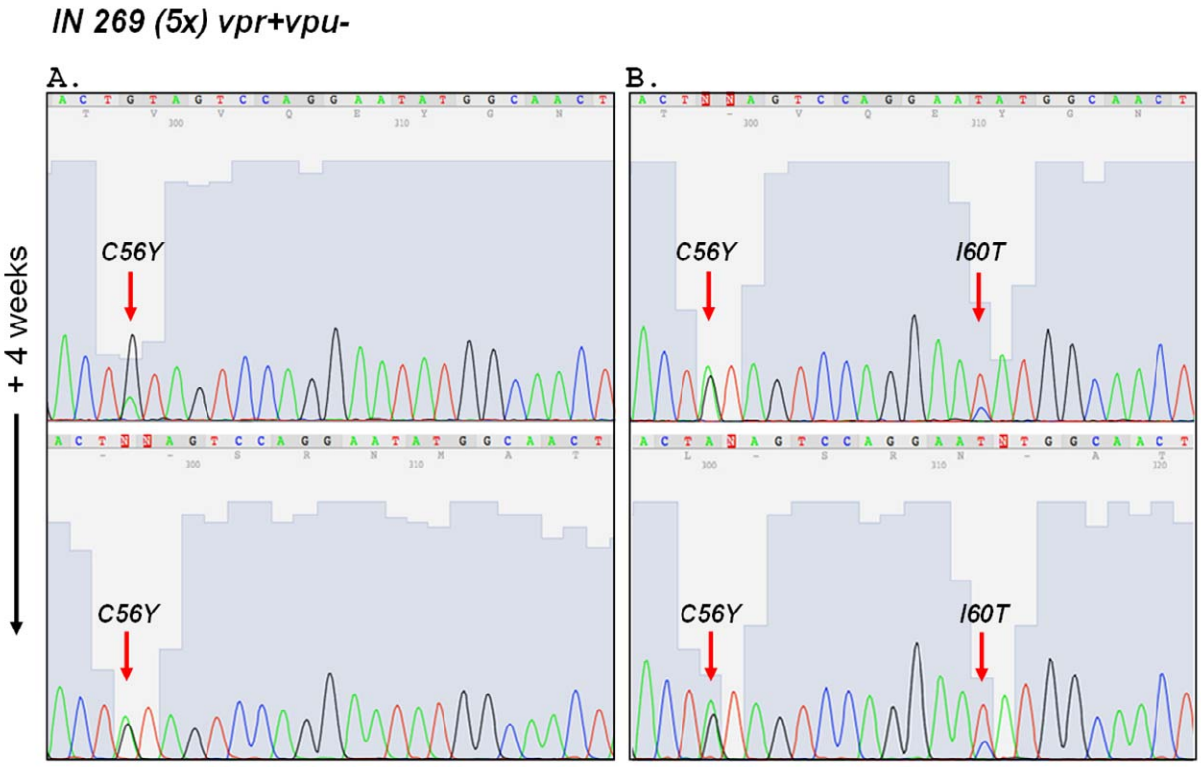


Figure 34. Outgrowth of revertant second site mutant viruses. The IN gene from viral DNA procured from infected cultures was PCR amplified and sequenced. A. and B. are derived from one infection that was split into two flasks, and followed in parallel. The C56Y mutation was likely present prior to the flasks being split. Flask B showed sign of a second mutation, I60T. The lower panels are a snapshot of the sequence of viral population taken four weeks after the top panels.

Results-

After four months in culture, robust syncytia formation was detected microscopically in cells infected with the Vpu- founder virus. This culture was then split into two flasks, designated 'A' and 'B'. These were monitored for sustained syncytia formation, and at various time points the IN region was amplified and sequenced. Two

such time points, taken four weeks apart, are shown in the chromatogram traces in Figure 34. The presence of one subpopulation could be readily detected in flask A, whereas two subpopulations were detected in flask B. These correspond to non-synonymous substitutions creating the missense mutations, C56Y and I60T. In an effort to detect the presence of less prevalent mutations, the PCR amplified IN region was subcloned back in the virus plasmid backbone, transformed to bacteria, and the individual clones screened. The list of mutations from individual clones is shown in Table 1 and the frequency with which they were obtained. As expected, the majority of the clones contain the predominant mutations C56Y and I60T, but other mutations were also revealed. Remarkably, despite the detection of a sudden enhanced rate of syncytia formation, the *vpr*- founder virus culture displayed no apparent heterogeneity within the IN coding region, Instead, a stop codon was detected at position 22 within the *vpu* gene and perhaps an ancillary mutation within the *tat* gene at its 3' splice site. It is not known if these two mutations are present in the same viral genomes, or represent two distinct populations. Although the significance of the *tat* mutation has yet to be investigated, the selection for the *vpu*- virus is indicative of the relevance of Vpu for viruses growing in CEM cells.

Table 1. Identification of potential revertant viruses. IN, Tat, Vpu, Nef regions and 2-LTR junctions from revertant viruses derived from Vpr+Vpu- or Vpr-Vpu+ cultures were PCR amplified, cloned into the parental viral plasmid, and sequenced. The non-synonymous mutations discovered are listed with the frequency with which they were detected in unique clones

IN 269 (5x) stop vpr+vpu- 2nd Site Revertants		
	Mutation	**Frequency**
Integrase	*C56Y*	*10*
	I60T	*4*
	I162T	*1*
	K14R C56Y	*1*
	C40R C56Y	*1*
	C56Y G140E	*1*
	C56Y Q164R	*1*
	C56Y Y194C	*1*
	C56Y Q148Q S153T	*1*
	I60T C65Y M154I	*1*
	E13K C56Y G82E E157K	*1*
Other Mutations	silent vpr F72F mutation/tat 3' splice site mutation	
	vpr stop opened (+12aa)/tat R7G	

IN 269 (5x) stop vpr-vpu+ 2nd Site Revertants		
	Mutation	**Frequency**
Integrase	*not detected*	*0*
Other Mutations	vpu W22 Stop	
	tat 3' splice site	

Analysis of Revertant Viruses:

Replication competency-

Present in the list of possible revertant viruses are several viruses containing multiple substitutions. These can be authentic revertant viruses, or potentially, these mutation combinations could be artifacts of recombinatorial events between distinct viral sequences that occur either during reverse transcription in the tissue culture dish, or during PCR amplification. We only tested the potential revertant mutations in the combinations as they were discovered, however. We intend to investigate each of these mutations in isolation in the near future. The viruses were tested for improved fitness in side-by-side comparisons with the founder IN 269 (5x) stop virus in the syncytia formation assay, and against WT virus, using a range of initial inoculums. The results are shown in Table 2.

Results-

The only mutations that measurably improve the fitness of the tail-less IN mutant in the syncytia formation assay are the C56Y and I60T mutations. These viruses were further characterized in efforts to understand how IN 269 was improved for replication competency by each of these additional mutational alterations.

Table 2. Testing the replication competency of possible revertant clones.

Mutant	Replication competency
C56Y	+++
I60T	+++
K14R C56Y	+++
C40R C56Y	-
C56Y G140E	-
C56Y Q164R	-
C56Y Y194C	-
C56Y Q148Q S153T	-
I60T C56Y M154I	-
E13K C56Y G82E E157K	-
269 5x (Parental)	+
WT	+++++

Characterization of C56Y and I60T Reversion Mutations:

Site-directed overlap PCR mutagenesis was employed to combine either C56Y or I60T with IN 269 and the resultant viruses (C56Y/IN 269 and I60T/IN 269) examined for the efficiency of 3' processing (8 hpi), the production of 2-LTR circles (24 hpi) and the relative efficiencies of the reconstituted mutants in the BSD transduction assay. The results are shown in Figure 35.

Results-

The results of the three assays are shown in Figure 35. In each case it is evident that both of the revertant mutations C56Y and I60T greatly enhance the fitness of the IN 269 truncated, tail-less virus. The basis for the rescue can be readily detected at the 3' P reaction.

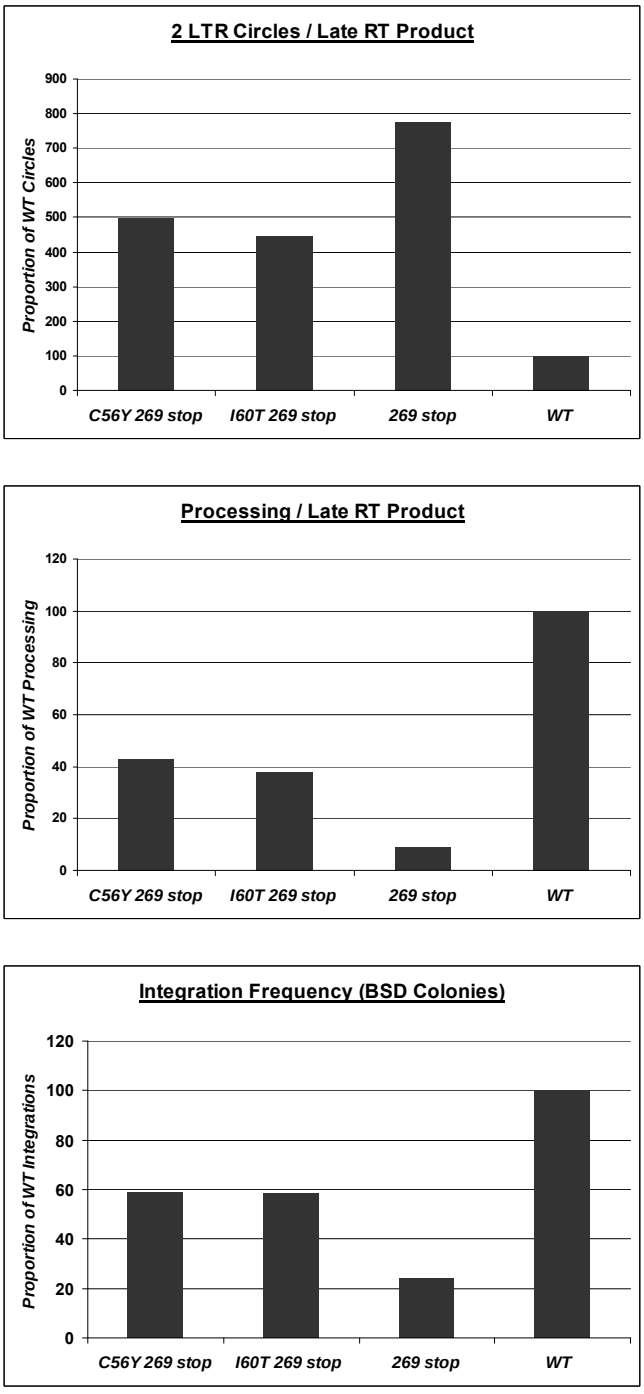


Figure 35. Rescue of tailless IN catalytic activity by second site revertant mutations.

Summary/Discussion-

We have successfully selected two intragenic second-site revertant viruses that have regained a significant level of fitness in the IN 269 'tailless' background. These mutations (C56Y and I60T) substantially improve the 3' processing efficiency of the parental, truncated IN 269 virus. This then ostensibly facilitates an increase in transduction efficiency, and an attendant diminution in the production of 2-LTR circles, a further indication of the improved catalytic function. Though there were fourteen mutations that were generated, only two, C56Y and I60T display observable improvement in replication competency.

The most remarkable discovery in this experiment was that the majority of the mutations that were generated mapped to the catalytic core of IN (Figure 36). Further validating their significance is the fact that these mutations were all clustered on the solvent-exposed face of a monomeric subunit in the dimeric crystal. The only non-CCD generated mutations were found in the NTD, and intriguingly one of them, K14R has been implicated in associating with LTR DNA during catalysis [132]. This mutation in isolation did not improve the fitness of the parental virus (IN 269), but in combination with the C56Y mutation facilitated enhanced replication kinetics, though not above the C56Y mutant alone. The role of the other detectable subpopulations that were detected outside of the IN reading frame are not known at this time. The *vpr* mutation that opened the reading frame (Table 1), extending the protein by 12 aa is provocative, but the overlapping substitution mutation in *tat* may be the more consequential mutation.

Likewise the detection of an alteration of the *tat* splice acceptor may have ramifications for replication. These mutants need to be tested in the future.

Though the *vpr*-*vpu*+ founder virus did not yield any compensatory mutations within the IN reading frame, the mutations that did accrue were nonetheless instructive. The loss of the *vpu* gene highlights the preferential mode of transmission within the CEM cell line. We also detected another mutation that overlaps with the highly conserved AG dinucleotide that controls the 3' splice site for *tat*. Given Tat's role in the transcriptional activation of the viral LTR, it is likely that an enhancement of viral transcription, comingling with the loss of Vpu function is sufficient to replicate a state in which virus production crosses the threshold required to produce syncytia in an efficient fashion. The impact of the *tat* splice acceptor mutation needs to be taking into consideration, if just to observe whether that *tat* spliced message is upregulated in these viruses.

A. **B.**

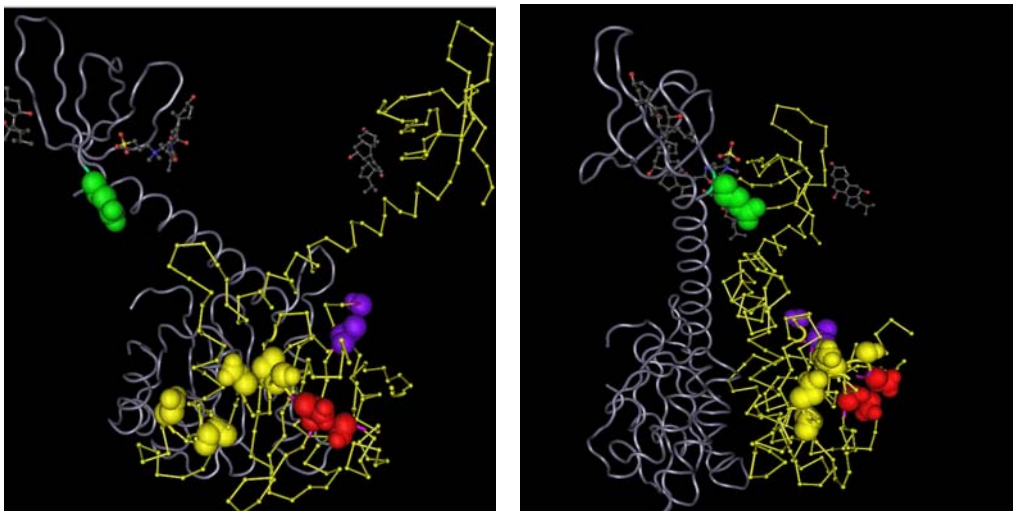


Figure 36. Location of mutated residues in revertant virus culture. The dimeric structure of IN, as solved by Stroud *et al*. [124], is depicted as a ball and stick model, with the CTDs extended away from each other, and the CCDs providing the dimeric interaction. The second-site revertant mutations are shown in space fill in one of the yellow monomer, the C56Y and I60T mutations shown in purple. Note that in the structure, Cys 56 was mutated to Ser. The catalytic triad is shown in red, and R269 on the CTD proposed to hold the viral DNA for the yellow monomer is shown in green. (B) is rotated 90° horizontally from the (A) view.

Development of Non-overlapping IN-VIF ORF Virus

With the exception of EIAV, which does not encode a recognizable *vif* gene, a universally-shared feature of the lentivirus genome is an overlap in the *pol(in)* and *vif* ORFs (Figure 37A and B). It is perplexing that HIV-1, so dependent on its genomic plasticity to elicit variants that evade immune responses, would adopt a genetic organization that encumbers the independent evolution of its gene products. The tail of IN, encompassed in the region of the overlap, is a relatively well-conserved region of the CTD, in part due to the restriction imposed by the overlapping N-terminus of Vif. Indeed, this region of Vif has been noted for its role in defining interactions with discrete APOBEC family members. We sought to create a virus in which the reading frames of *pol(in)* and *vif* are distinct, thus providing the flexibility to manipulate the IN tail without simultaneous and unintentional Vif perturbation. The ultimate goal is to use these viruses to evaluate the requirements of the IN tail during infection of primary CD4+ T cells and macrophages, natural targets of HIV-1. Since these cells are Vif non-permissive, this virus would be a useful tool that would allow manipulation of the IN tail without disturbing the integrity of the *vif* gene or its product.

We created three related constructs that separated the *pol(in)* and *vif* reading frames (Figure 37C). In each, the native nucleotide sequence of the IN tail, starting with D270, is altered such that the third nucleotide of each codon triplet is replaced with a synonymous base pair that maintains the natural IN amino acid sequence. The alteration of the IN tail sequence in these constructs serves two purposes: it removes the natural *vif*

ATG initiation codon from the IN reading frame, and it avoids the juxtaposition of two identical sequences, a direct repeat that would likely be unstable with a propensity to be removed by recombination. This original construct as just described is designated IN-VIF. Two iterations of this original construct were made, that may enhance the efficiency of Vif protein expression. The first, called IN-VIF [IN(M275V)], substitutes the methionine residue at position 275 in IN with valine, a naturally occurring polymorphism found at this position.. In HIV-1, Vif is produced by alternative splicing of the viral mRNA; the 3' splice site controlling its expression lies just upstream of the natural initiation codon of the *vif* gene. In our construct, this modification of IN M275V residue makes the ATG initiation codon of *vif* the first ATG encountered by the ribosome on the spliced transcript, thus increasing the efficiency of its translation. The second construct iteration is called IN-VIF [VIF (KOZAK)]. In this construct, the four bases preceding the vif initiation codon are altered to a canonical Kozak sequence (CACCATGG), which would likewise enhance the efficiency of translation initiation at this ATG codon, underlined in the sequence above.

We tested the above viruses for replication competency first in immortalized cell lines. We selected both a permissive (CEM), and non-permissive (MT-2) cell line for these initial experiments, observed the infected cell for the formation of syncytia, which is indicative of productive infection (Figure 37D).

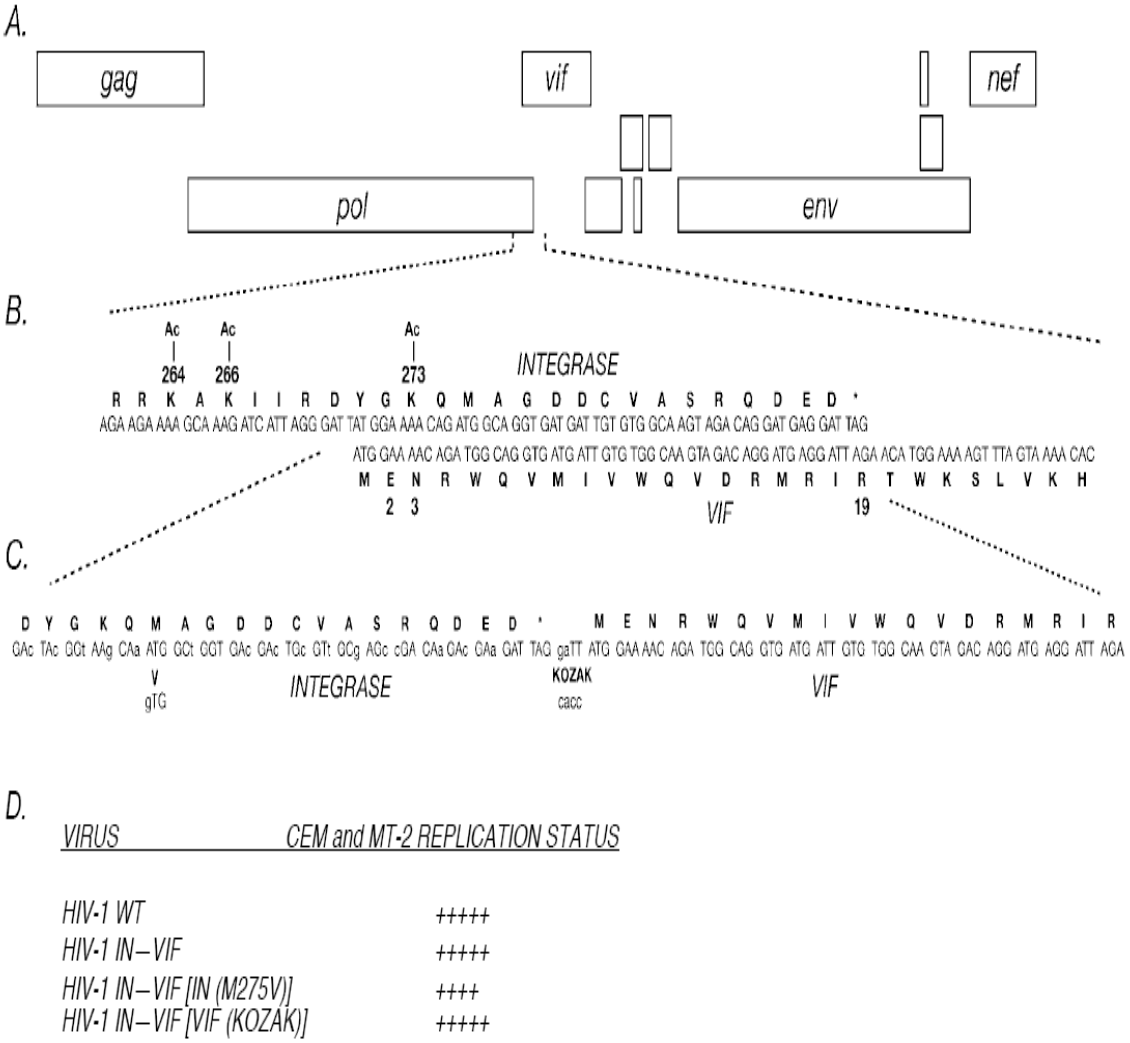


Figure 37. Separation of the Pol and Vif reading frames allows for the independent study of residues at the C-terminus of IN. A) Reading frames of HIV-1 depicted as blocks. (B) Expanded view of the overlap sequence between IN and Vif. C) Mutagenic separation of the reading frames. gTG (M275V) and cacc (KOZAK) are alternative mutations constructed to facilitate Vif expression in IN-Vif virus. D) Results of syncytia formation assay in Vif-permissive (CEM) and Vif-nonpermissive cells (MT-2).

Results-

The results are presented in Figure 37D. The IN-VIF constructs each created syncytia at a comparable rate with the WT overlapping reading frame virus. Only the [IN(M275V)] virus showed a slight delay in syncytia formation.

Investigation of Vif protein production:

Replication in the non-permissive MT-2 cell line indicated that our viral constructs produced sufficient amounts of IN and Vif proteins that are required for productive infection. However, the expression of a subset of HIV-1 genes, that include Vif, is dependent on an optimized pattern of alternative splicing of the full-length viral mRNA. Manipulating the viral genome has the potential to disrupt the delicate balance of splicing, which can have global consequences on the expression of viral proteins. The region that we manipulated lies just downstream of the pair of 3' and 5' splice sites that facilitate the production of the *vif*-encoded spliced mRNA when the 5' slice site is suppressed. The organization of these sites in relation to the *pol(in)* and *vif* reading frames are shown in Figure 38. This organization of splice sites is a common feature in the HIV-1 genome, and confers a level of control over the production of the late viral gene transcripts. For example, at some frequency a sub-optimal downstream viral 5' splice site can be over-utilized, thus bypassing the downstream viral ORF coordinated by utilization of the 3'- and suppression of the 5'-splice site, Conceivably, the over- or under-activity of either splice site would affect not only production of that protein, but that of downstream proteins as well. We investigated the levels of Vif protein production

in the IN-VIF viruses to determine if Vif was being produced in natural amounts. Cells were transfected with the various constructs, lysed, and then probed for vif protein content. The results are shown in Figure 39.

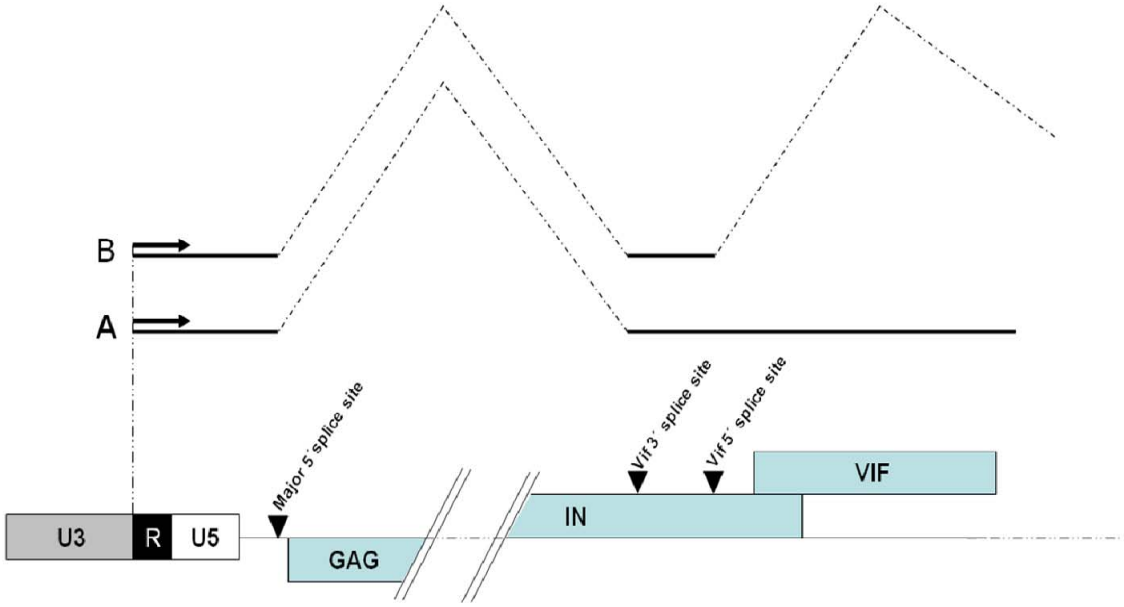


Figure 38. Splice site organization for HIV-1 Vif reading frame. Transcription of the viral genome starts at the boundary of the R region in the 5' LTR. Splicing occurs between the HIV-1 Major Splice Donor (MSD) and downstream splice sites that lie upstream of the all HIV-1 regulatory proteins, and for Env. At some rate, a 5' splice site downstream of the 3' splice site is utilized, bypassing the downstream ORF of the protein.

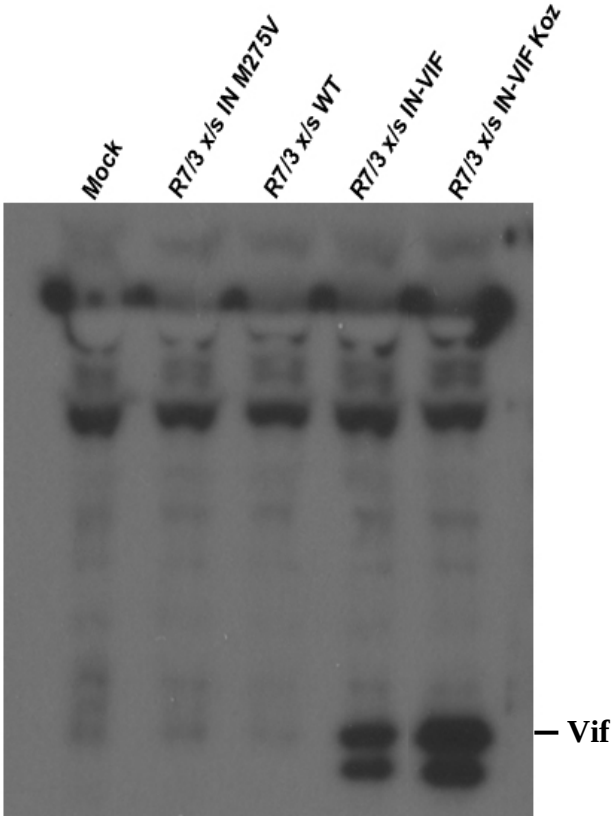


Figure 39. Investigation of Vif protein production in IN-VIF viruses. 293T cells were transfected with the various constructs, lysed, and the cellular extracts were probed for Vif protein content by Western blot analysis. The double bands observed for Vif is attributed to the utilization of a downstream ATG codon in the Vif ORF that creates a slightly shorter version of the Vif protein.

Results-

As shown in figure 39, whereas the Vif protein expressed from the WT plasmid is undetectable over the background, the separated reading frame viruses IN-VIF and IN-VIF KOZ express an overabundance of Vif. The basis for this disruption might be attributable to disregulation of the *vif* 3' splice site. This was investigated.

vif Gene Splicing:

We next tested the levels of the *vif* splice product in transfections of our three viral constructs. Total RNA was harvested, reverse transcribed and then subjected to PCR analysis with a primer pair specific to *vif* or to the downstream *vpr* splice mRNA species acting as a control for these experiments (Figure 40). The results show that compared to the WT virus, the IN-VIF and IN-VIF [IN(M275V)] viruses appear to be making significantly more *vif* specific message, so much so that the *vpr*-specific primers are preferentially detecting *vif* message over the *vpr* message. The IN-VIF [VIF (KOZAK)] virus on the other hand has a more typical splicing pattern, and most closely resembles the splicing pattern seen in the WT.

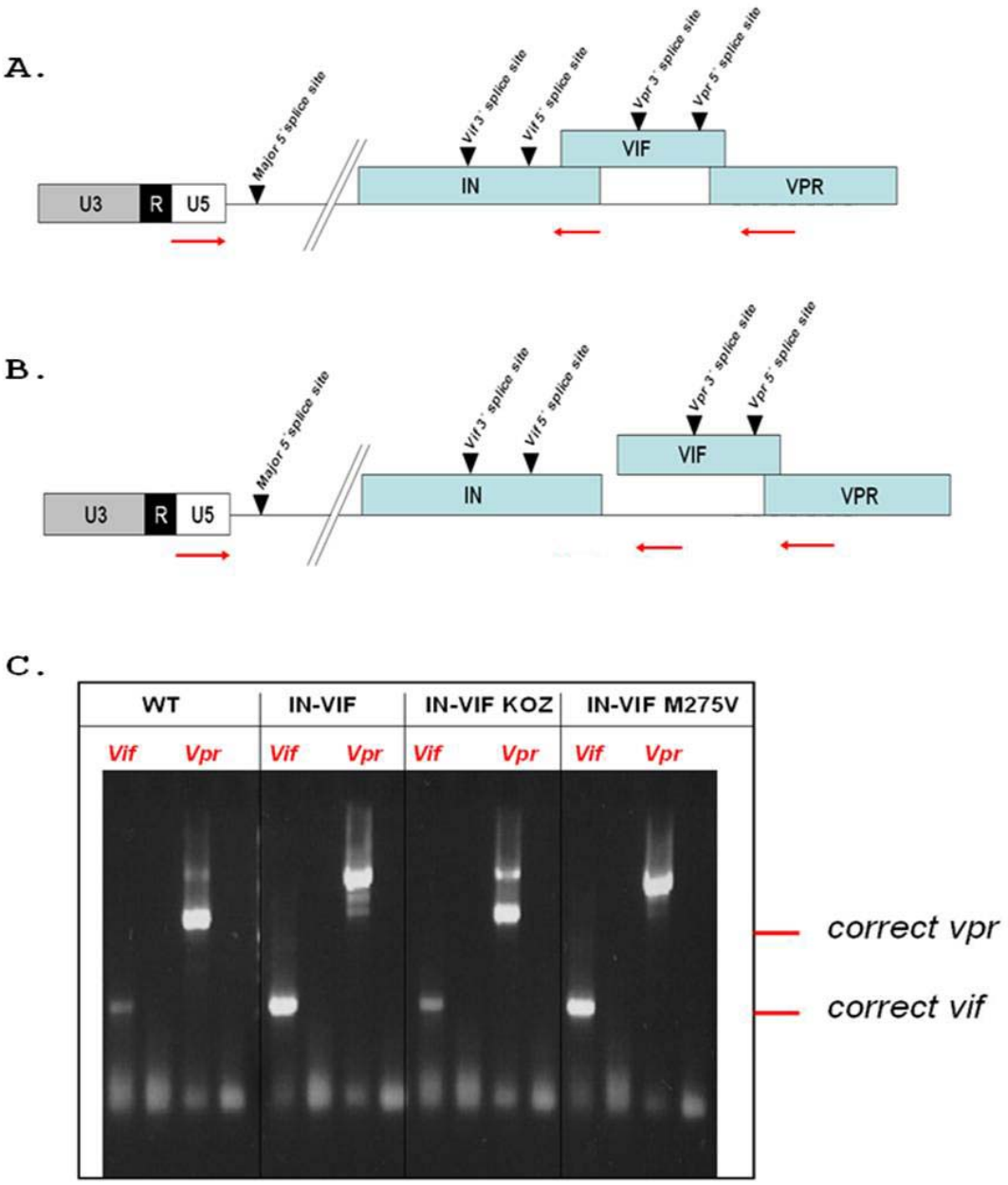


Figure 40. Qualitative PCR of Vif and Vpr spliced products in IN-Vif separate ORF viruses. Cells transfected with each virus construct were harvested for RNA, which was reverse transcribed, and Vif and Vpr specific primers were used to detect the levels of Vif and Vpr spliced messages.

Primary Cell Infection:

These constructs were used in a preliminary experiment to see if they would infect primary cells. The cells used for this experiment were CD4+T memory cells, obtained by negative selection over a magnetic column, and activated with phytohemagglutinin (PHA) prior to infection. The cultures were followed for several days, and culture supernatant was assayed for the presence of p24 antigen at regular intervals. The results are shown in Figure 41.

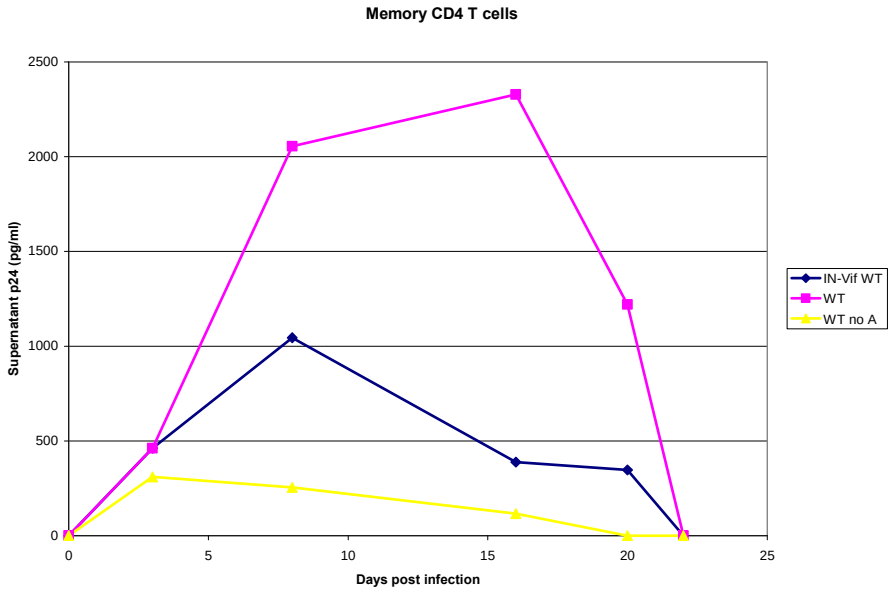


Figure 41. Replication of IN-VIF virus in primary cells. Memory CD4+ T cells were stimulated with PHA one day prior to infection, and challenged with WT virus, and IN-VIF virus. As a control, unstimulated CD4+ cells (WT no A) were infected with the WT virus. Viral replication was followed by measuring p24 CA antigen released into the supernatant. IN-VIF Virus-Dark blue; WT virus-Pink; Unstimulated cell infection-Yellow.

Discussion-

The IN-VIF separated reading frame viruses show great promise for use as a tool for manipulating of the IN tail. Significantly, they are replication competent in primary cell culture, and though the kinetics are somewhat attenuated, this signifies that the virus can tolerate this genetic manipulation on at least the short term required for the planned experiments. For our purposes, it may be necessary to initially passage the virus through primary cells or a Vif non-permissive cell line to allow the virus to fine-tune itself for enhanced fitness. The consequences of the massive Vif overproduction are not known, but if it is detrimental it is possible that the serial passaging would repair the problem. Furthermore, such experimentation will reveal whether several rounds of replication of our manipulated genomic architecture can be tolerated by the virus.

CHAPTER 4- DISCUSSION

Unique biochemical activities are often compartmentalized to distinct domains in a single polypeptide chain, enabling an enzyme to orchestrate a particular biological process. The elucidation of the domain repertoire of HIV-1 IN engendered the application of such reductionist analyses, with attempts by many researchers to reconcile inherent biochemical activities of isolated domains with the obligatory requirements of the integration reaction. It quickly became clear, however, that when applied to IN this simplistic allocation of specific activities was inadequate. Instead, a picture has emerged in which each domain is implicated in contributing cooperatively to the requisite DNA binding and protein-protein interactions in the minimally functional integration complex, or intasome. The architecture of the HIV-1 IN-DNA complex is still a matter of conjecture, and awaits the resolution of the structure of HIV-1 IN complexed with its viral and target DNA substrates. Thus, the critical contributions of each domain can only be inferred empirically from other sources.

The above considerations however fail to take into account the secondary roles of HIV-1 in the viral life cycle. In the context of GagPol, IN influences the assembly process by providing an integral but as yet, ill defined role in the regulation of precursor processing. In addition, during the infection of target cells, IN contributes to the efficiency of reverse transcription. This aspect of IN biology has been a source of frustration in the field, for conventional dissection of protein function by directed mutagenesis studies must always be considered with the potential simultaneous disruption of primary and secondary functions in mind. Because of this, *in vitro* studies of

IN functions, though critically insightful for studying facets of IN catalytic activities, provides only a small view of the breadth of the activities attributable to IN *in vivo*. This thesis describes a survey of IN CTD truncations aimed at deciphering the role of the IN C-terminal tail, and by extension, the involvement of the entire CTD in the context of viral infection. Most queries of mutant function I describe examine events at positions within the life cycle in which IN is known to be an active participant. This includes post-translation of the Gag and Gag-Pol precursor polyproteins, virion maturation and its egress in the producer cell and those steps in the target cell that lead to the establishment of the proviral state. Appraisal of replication defects associated with IN functionality were dissected using a combination of novel and conventional methodologies, aimed at uncovering the specific activities lost by specific CTD truncation. This information provides a framework for follow-up focused investigations of IN activities *in vitro,* settings that can suitably eliminate confounding variables from consideration.

During the preparation of this dissertation, the highly anticipated crystal structure of a retroviral integrase in complex with its substrate DNA was solved to high resolution [199]. Although the structure is derived from the distantly-related prototypical foamy virus (PFV) IN, the potent susceptibility of the enzyme to the current INSTI drugs makes this achievement particularly relevant to the HIV-1 IN field. From the standpoint of previous theoretical models, the data divulges a wholly unanticipated arrangement of individual subunit domains within monomeric IN and subunit arrangement within the tetrameric intasome. Nonetheless and in agreement with an inferred consensus gleaned from the plethora of HIV-1 IN studies, each domain in the PFV IN structure contributes

to both protein-DNA and protein-protein contacts. Though it is presently unknown whether the extrapolation of the structural arrangements of the PFV intasome onto that of HIV-1 is correct in all of its manifestations, wherever possible, our results are interpreted below in light of this new structural model.

Three-dimensional information for the HIV-1 IN CTD abruptly ends at Asp 270, the ultimate residue of an interwoven mesh of 5 interconnected β-sheets spanning 48 amino acid residues. The overall topology assumes an SH3-like fold, to which is appended the final 18 residues constituting the disordered tail. PFV IN likewise contains a disordered tail of equal length, recalcitrant to structural resolution apparently due to a high degree of flexibility. One concern of the potential consequence of sequential removal of residues past the 270 boundary and into the structured region is the disruption of intimate inter-β-sheet interactions that sustain the three-dimensional structural integrity of the SH3-like element and the CTD. Some of our observations certainly hint that such a situation might be occurring by truncating past IN 268. Although IN 268 is a replication-incompetent Class II viral mutant, its behavior in several assays reflects the preservation of a nominal IN activity that is lacking with further truncation. Indeed, removal of amino acids past Ile 268 (i.e. IN 267, 266, 265, 260) creates a mutant class that behaves remarkably similar to a mutant that is completely devoid of the CTD (IN 212), in terms of intracellular processing, virion Gag-Pol packaging and in their ability to complement a CCD catalytic mutant (D116A). Furthermore, the eGFP-IN 267 fusion protein creates large cytoplasmic aggregates (Figure 32), a behavior described for misfolded proteins. Although IN 267 and IN 268 both terminate in hydrophobic Ile, the eGFP-IN 268 fusion

protein appears to function normally by interacting with LEDGF, and apparently not disposed to aggregation. This may hint that the Ile 268 residue is participating in its appropriate context. Paradoxically, the eGFP-IN 266 fusion protein terminating in Lys266 does not aggregate, a result that may indicate that if solvent exposed, Lys266 may favor solubility of the protein. Taken together, the intermediate behavior of the IN 268 viral mutant, its length just one amino acid short of replication-competent IN 269, suggests that Ile268 could represent the final "nail" holding and maintaining at least a portion of the SH3 fold. In fact, the NMR structural data for HIV-1 IN CTD in solution reveals that Ile268 is the last residue of the fifth and final β-sheet (residues 265-268) of the SH3 fold. Close inspection of the NMR structure indicates that Ile268 is in intimate contact with Phe226 in a proximal β-sheet. This interaction contributes to the establishment of a bridge of hydrophobic interactions encompassing the three sheets (β 1, β 2 and β 5) across one face of the fold (Figure 42).

If our assertion of disparities in CTD integrity is accurate, then the complementation studies of the IN truncation panel support current mechanistic models for the HIV-1 integration reaction. Numerous studies support a model in which each viral cDNA end is processed by an IN dimer that likely assumes the orientation discovered in the HIV-1 IN crystal structure, depicted in Figure 36. The authenticity of the dimer interface, created by vast CCD-CCD contacts, is supported by the presence of this identical interaction in the PFV intasome structure. The physical separation of the CTD domains is due to an asymmetry between the two monomers created by a kink in the CCD-CTD α-helical linker in one subunit. The viral cDNA is proposed to interact with

the IN dimer in such a way that the end of the DNA molecule lies in the catalytic core of one subunit, while the DNA molecule extends away along the outer aspect of that subunit

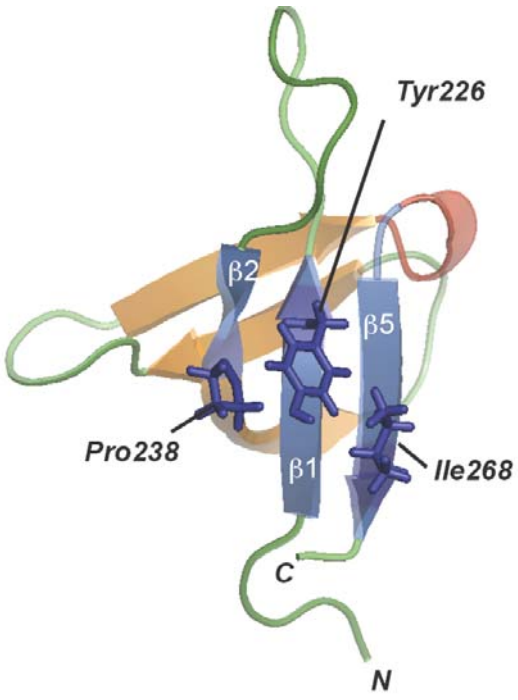
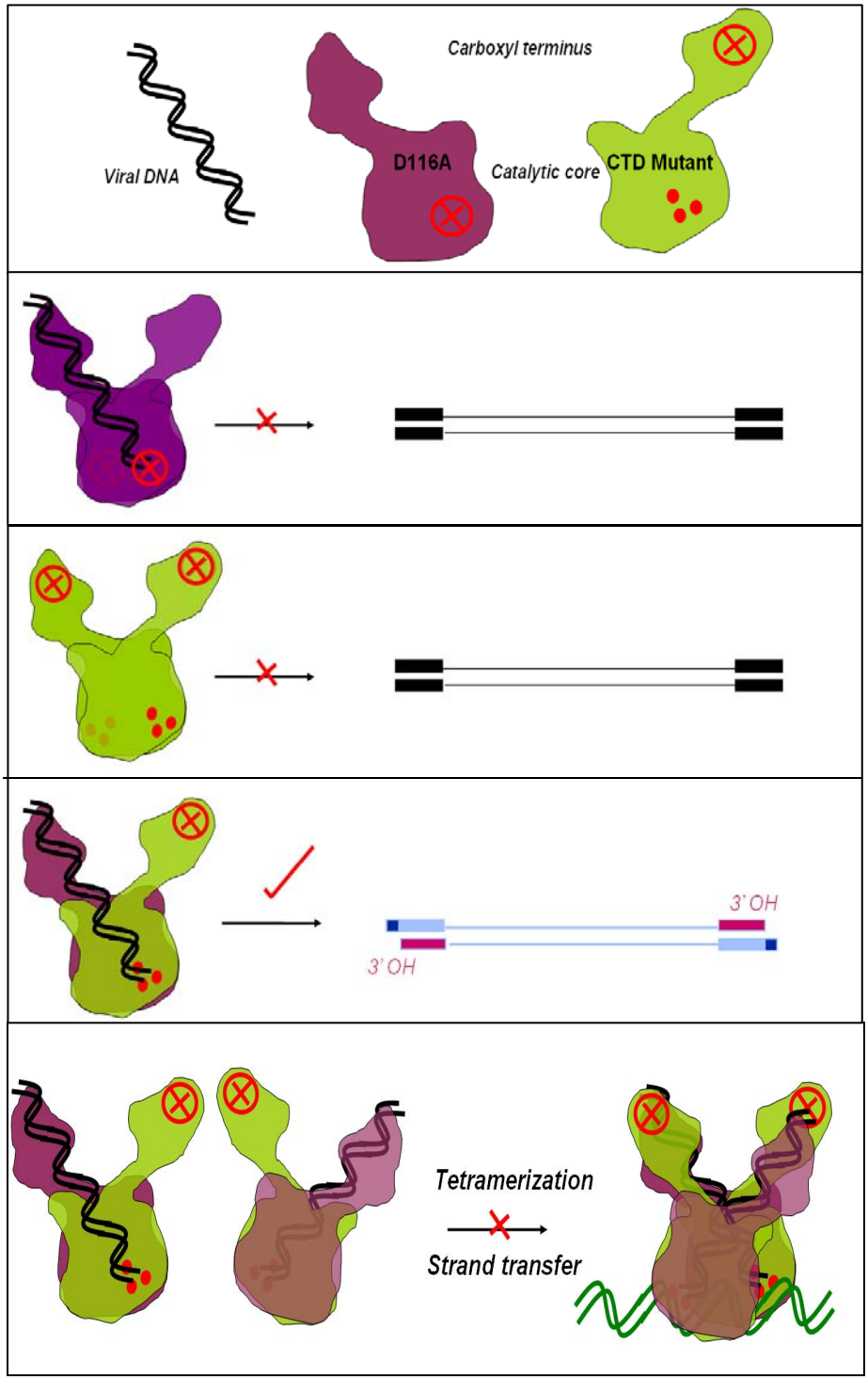


Figure 42. Hydrophobic interactions in the SH3 fold of the IN CTD. The SH-3 fold of the IN CTD displayed with beta sheets 1, 2a and 5 in blue, and 2b, 3 and 5 in ochre. The side chains of the three hydrophobic residues making putative interaction are shown in stick model.

to then establish an association with the CTD of the second monomer. In this way, the CTD of one subunit stabilizes the DNA, while the opposed subunit conducts the enzymology that directs 3' processing. This premise forms the basis for our model (Figure 43) proposed to account for our observations of the IN truncation complementation. IN 267, for example, contains a catalytically-active CCD, which according to the model can be supplied *in trans* to the IN D116A mutant, which reciprocally supplies an intact CTD, thus satisfying the requirements for processing. If

what we assert is accurate, then although IN 267 has an improperly folded CTD, processing can proceed because of the particular arrangement of IN molecules during processing. This complementation is what was observed in both IN 266 and IN 267 viral complementation experiments (Figure 30B). In contrast, strand transfer is coordinated by at least a tetrameric assembly of HIV-1 IN monomeric/dimeric units and is thought to be contingent upon completion of the previous catalytic step (3' processing). Indeed, the tetrameric assembly of PFV IN on its pre-processed DNA substrate used for crystallization studies provides very strong evidence for such a premise. Unlike the dimeric IN unit, which relies solely on CCD-CCD interactions, it is believed that the HIV-1 IN tetramer, like that of PFV IN, is stabilized in part by protein-protein interactions facilitated by the CTDs from distinct dimeric units. Thus, despite achieving relatively significant levels of processed DNA ends, loss of the essential contacts required to create the tetrameric strand transfer complex may preclude the second integration step. This is reflected experimentally in the relatively poor integration rate of the hybrid IN 267/IN D116A virus, in spite of rather efficient processing at each of the viral DNA ends (Figure 30 A and B). IN activity however extends beyond integration, and the unique behavior of the IN 268 was evident also in its intracellular processing profile. Indeed, the addition of Ile 268 (IN 268) did not fully rescue the processing defects of the Class II IN truncation mutants, but it did significantly alter the processing of the Gag-Pol polyprotein precursor, establishing a semblance of the WT pattern. Our data suggest that the altered pattern of precursor processing observed for the Class II mutants was precipitated by disregulated viral protease activity. Though the mechanisms

Figure 43. Model for the complementation of CTD truncation mutants. IN D116A possesses no catalytic activity, but is functional for all other aspects of the viral life cycle. Significantly, it possesses an intact CTD domain, which is the deterrent for the IN CTD truncations. In the first iteration, if two D116A monomers associate, 3' processing does not occur due to both monomers lacking a catalytic site. The IN truncations, for example IN 267, possesses an intact catalytic site, but the disrupted CTD can no longer function to stabilize the DNA in the catalytic core of its opposing subunit. When IN D116A and IN 267 associate however, the intact CTD of the IN D116A subunit can successfully supply *in trans* that activity that is inherently missing from IN 267, and so processing can proceed. The strand transfer reaction however is thought to require at least a tetrameric arrangement of IN molecules held together partly by CTD interacts. The disrupted CTD of IN 267 does allow some strand transfer to occur in the context of the hybrid IN D116A virus, but not as efficiently as other viruses that support a similar level of 3' processing (IN 269, Figure 30)


Carboxyl terminus
Viral DNA
D116A
Catalytic core
CTD Mutant
3' OH
3' OH
Tetramerization
Strand transfer

underlying viral protease activation are unclear, studies indicate that dimerization of the Gag-Pol precursor is a necessary prelude, since the active site of the viral protease is comprised of two individual protease subunits [49]. *In vitro* studies indicate that the initial cut made in the Gag-Pol precursor by the activated protease occurs at a predestined cleavage site in the precursor dimer, usually at the p2 spacer/NC junction [189]. The subsequent precursor processing ensues in an ordered and temporally staggered fashion until full processing is accomplished. A perhaps significant observation is that the PR embedded in Gag-Pol can efficiently cleave only a limited set of protease sites, whereas the mature PR is capable of efficient cleavage at all sites [189].

It has been noted previously that limited IN CTD truncation or its complete ablation leads to premature precursor processing, but the mechanistic details have not been fully explored. While one possible explanation is that the CTD in the Gag-Pol context could deter precursor dimerization, we observe a large quantity of intracellular, partially processed products in the WT virus, indicative of precursor dimerization leading to initial PR processing. Naturally, the tissue culture system in which we are studying this phenomenon is artificial; *in vivo* infections likely cannot support the extent of viral gene production that is possible from a transfection of highly competent HEK 293T cells. Consequently, the levels of Gag-Pol in a natural setting would not likely achieve levels high enough to possibly accelerate intracellular dimerization. An alternative mechanism for role of the CTD in processing is that it may influence the selection of initial cleavage sites of the embedded PR concomitant with Gag-Pol dimerization, thus directing a pattern of cleavages that precludes the production of mature protease at early time points. The

Western blot in Figure 27 seems to indicate this very proposition, as there is a clear difference in the cleavage pattern of IN 266 and 267 in the presence of Ritonavir. Studies have shown that the embedded PR is up to 10,000x less sensitive to Ritonavir than the mature PR, and this phenomenon has been used to dissect the initial cuts made by PR concomitant with precursor dimerization [49]. One can envisage a number of mechanisms by which this could occur, including physical occlusion of protease sites by the IN CTD, or by a direct interaction of the CTD with the PR domain, or alternatively with the RT domain, a region of Gag-Pol implicated in facilitating the dimerization of Gag-Pol precursors [47, 171]. Alternatively, the CTD could recruit cellular proteins that could influence PR activity. One such candidate is Integrase Interactor 1 (INI1/hSNF5), an identified cellular binding partner that binds IN within the context of Gag-Pol and is packaged into virions [174]. The intriguing aspect of this interaction is that if it is abrogated, the viral infectivity is blocked at the late stage of replication, during assembly [174]. The CTD of IN has been implicated in facilitating the binding of several other cellular proteins, and future studies of the interaction capabilities of our IN CTD truncations with known cellular binding partners may prove fruitful.

IN 268 is nonetheless a Class II mutant, unable to support RT activity. It is also apparent that the defect in its CTD also blocks its ability to conduct the integration reaction, based on our experiment with IN 268 and the Class I IN 270 experiment. However, we find that IN 269 to be replication competent, indicating that at this CTD length, IN is possessed of not only its primary integration activities, but also by necessity, all of its ancillary functions. Indeed our data demonstrate that IN 269 virus exhibits

accurate intracellular processing and efficient reverse transcription, abilities both lacking in the IN 268 mutant. The arginine residue at 269 has however not been implicated as a particularly critical CTD residue. In the context of the full-length protein, an R269A mutant is replication competent, albeit with significantly delayed kinetics, and was classified tentatively as Class I to indicate that the defect was specific for integration [2]. So, in light of the significance of the R269 residue for IN catalysis, and not for reverse transcription, our data suggests that the sudden gain of secondary IN functions for IN 269 for both the precision of $Pr55^{Gag}$ and $Pr160^{Gag\text{-}Pol}$ processing and RT activity is not due to the sudden acquisition of specific interactions *per se*, but is more likely explained by an enhanced stabilization of the SH3 fold that is induced by inclusion of this residue. Indeed, recent reports describing a functional IN/RT interaction mapping to the IN CTD [109, 110, 148, 149] and plasma resonance analyses [148] identifying specific CTD residues mediating the RT interaction did not include R269 among the specifically interacting residues, highlighting an indirect role of this residue in promoting reverse transcription.

With the gain of normal, albeit diminished function in IN 269, and our assertion that the R269 residue might enhance CTD stability in the context of the truncated protein, we were surprised to find that the addition of the subsequent one to three residues results in a loss of replication competency (mutants IN 270, 271 and 272). The loss of replication status of these particular mutants is not likely due to the misfolding of the SH3 fold, as the late RT and complementation experiments suggest that it is probably intact. These IN mutants advance through the preceding viral life cycle stages with minimal defect but are defective specifically for catalysis, demonstrated by a lack of viral

3’ end processing. Our data thus suggests that a correctly folded and intact SH3 fold sufficiently facilitates the secondary IN functions (precision of processing, RT activity), but more is required for integration itself. What component of IN catalytic activity necessary for the integration reaction itself is lost by adding one, two or three additional residues to IN 269? Given that 3´ processing is affected, it is feasible that the appendage of the one, two or three residues corresponding to IN 270, 271 and 272 to the IN 269 mutant curtails the CTD-viral DNA binding required for stabilization and positioning of the DNA end in the active site of the opposing IN subunit in the integrase dimer. The addition of a single residue at 273, an invariant residue subject to both acetylation [3, 4] and possibly ubiquitination [153], rescues infectivity. Though it has been suggested that the K273 plays significant roles in integrase dynamics, including specific DNA binding, substitution of alanine (K273A Stop) or arginine (K273R Stop) instead of the natural lysine residue to the truncated protein is still sufficient to rescue integration activity to high levels (data not shown). Further biochemical analysis of these Class I truncations is required to determine the specific activity lost by addition of these additional residues to IN 269.

The function of the tail is still unknown after this work, but it is clear that it enhances both secondary and primary IN functions. For the precision of precursor processing, the contribution of the tail is clearly evident when comparing IN 282 and IN 288 (full-length) (Figure 26, lanes 16 and 17 and Figure 28, lanes 18 and 19). There was however, no indication that the tail enhanced the efficiency of RT activity (Figure 18). For primary IN function, the contribution of the tail is evident during the first step of

integration, 3' processing (Figure 24), with a gradual improvement of processing efficiency with incremental lengthening of the tail. From our data however, we cannot determine why this should be so. If the IN dimer model for the processing reaction is accurate, then the tail might serve to augment the affinity of the CTD for the viral DNA, making it better suited to lock the DNA in position for 3' processing at the distal end of the DNA molecule. Too short a tail (IN 270, 271, 272) is detrimental, perhaps a hindrance for productive DNA binding. A longer tail then may be increasingly flexible and capable of holding and locking viral DNA in place.

Selection for second-site revertants evolution experiments identified two mutations that significantly increased the integration efficiency of the tail-less mutant (C56Y and I60T), both of which, when using conventional domain boundary designations, map to the CCD (aa 50-212). Close inspection of the solved HIV-1 IN NTD/CCD two-domain structure however reveals that the precise location of Cys 56 is within an unstructured linker sequence between the NTD and CCD, whereas Ile 60 is at the boundary of the linker and the CCD, representing the initial residue of the first structural element within the CCD. Comparisons of the related residues in PFV IN reveal a provocative role of these particular residues in the assembled IN-DNA complex. The PFV intasome was assembled with a preprocessed DNA substrate, and intriguingly PFV IN Ile 112, the residue equivalent of HIV-1 IN C56 from structure-based alignment of the two proteins, comes into intimate contact with the thymine base of the fifth nucleotide in the unprocessed DNA strand. This nucleotide position is the first in the DNA substrate to conform to a double-stranded structure; the terminal five nucleotides of the unprocessed

strand are single-stranded and bent away from the complementary strand. Ile112 may thus play thus be involved in DNA binding and positioning in the active site before or during processing, and/or facilitating distortion of the DNA helix, which in combination with DNA end fraying is required for 3' processing [200, 201]. It maybe premature to make direct comparisons of the two enzymes, especially taking into consideration the differing lengths of the NTD-CCD linkers in the two proteins, but such a mechanism of enhanced binding or distortion of the viral DNA in the active site may explain why the C56Y mutant can rescue a purported DNA binding attenuation by the truncated IN 269 CTD.

In spite of lingering uncertainty of the actual mechanism by which the tail contributes to the integration reaction, it is apparent that the activity it exerts is most essential at 3' processing, and not for the subsequent strand transfer. Indeed, for the tail-less IN 269 mutant, and for the panel of truncations as a whole, the limiting step for transduction ability is 3' processing. Once the ends are processed however, the viral DNA is expeditiously and very efficiently integrated, indicating that strand transfer is not affected by the absence of the tail. This assertion is supported by the value of the integration rates of the tail-less virus, relative to the 3' processing levels (Table 3). This analysis indicates also that the second-site compensatory mutation improves the replication dynamics of the virus specifically for 3' processing. The subsequent strand transfer step actually appears to proceed somewhat less efficiently, when compared to the parental 269 stop virus (Table 3).

Table 3. Efficiency of tail-less IN viruses to complete strand transfer after processing viral DNA ends.

MUTANT	*LATE RT*	*3'Processing/late RT (%WT)*	*Frequency of Integration (%WT)*	*Freq Int/3' processing (%WT)*
C56Y 269 stop	**117.2**	**42.9**	**58.6**	**136.6**
I60T 269 stop	**129.7**	**38.0**	**58.2**	**153.2**
269 stop	**307.8**	**8.9**	**24.2**	**273.0**
WT	**100**	**100**	**100**	**100**

These considerations serve to further solidify the important role of the tail during 3' processing. Unfortunately, the solved intasome structure of the PFV does not divulge any pertinent information regarding the subunit arrangement of IN for 3' processing, since the complex was assembled with preprocessed DNA substrate, and as such is primed for strand transfer. In the PFV IN-DNA complex, a pair of asymmetric IN dimers come together to form the tetrameric complex. The inner subunits of the tetramer contribute all of the contacts involved in DNA binding and tetramerization, whereas the outer subunits, of which only the CCDs could be resolved, are postulated to contribute a supporting function. In current models of the HIV-1 integration, it is these outside subunits that provide the necessary CTD contacts on the viral DNA for the 3' processing reaction, in which the functional catalytic unit of IN is the dimer form. In the PFV structure, the NTD and CTD structures were not resolved in the outer subunits of the tetramer, and the apparent lack of contribution of these domains from the outer subunits

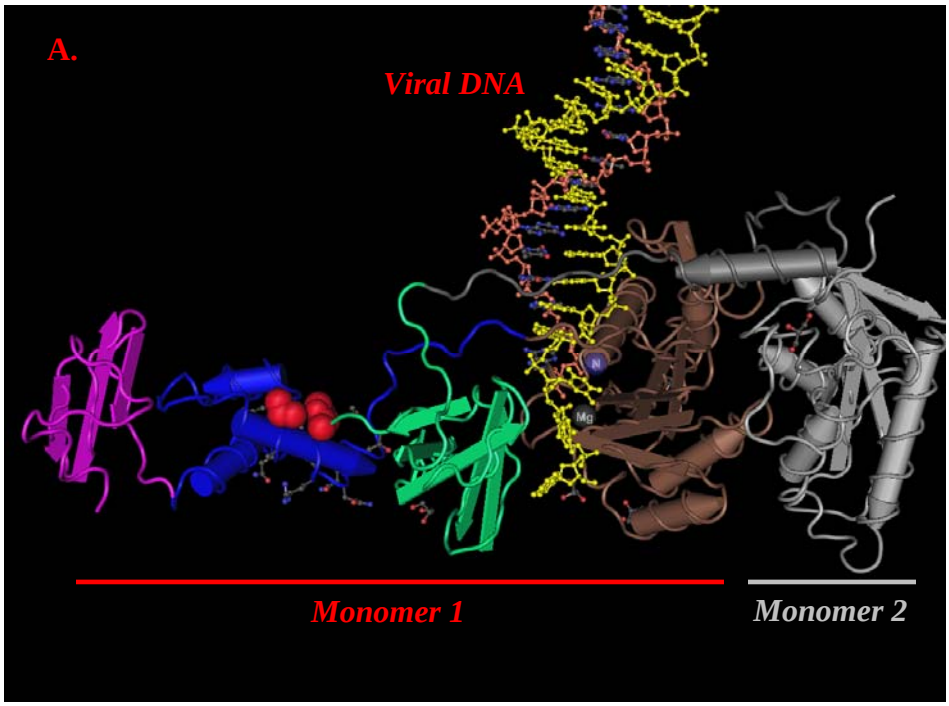
to the strand transfer complex is further proof of a function of the IN tail that is useful for processing, but subsequently unnecessary for strand transfer.

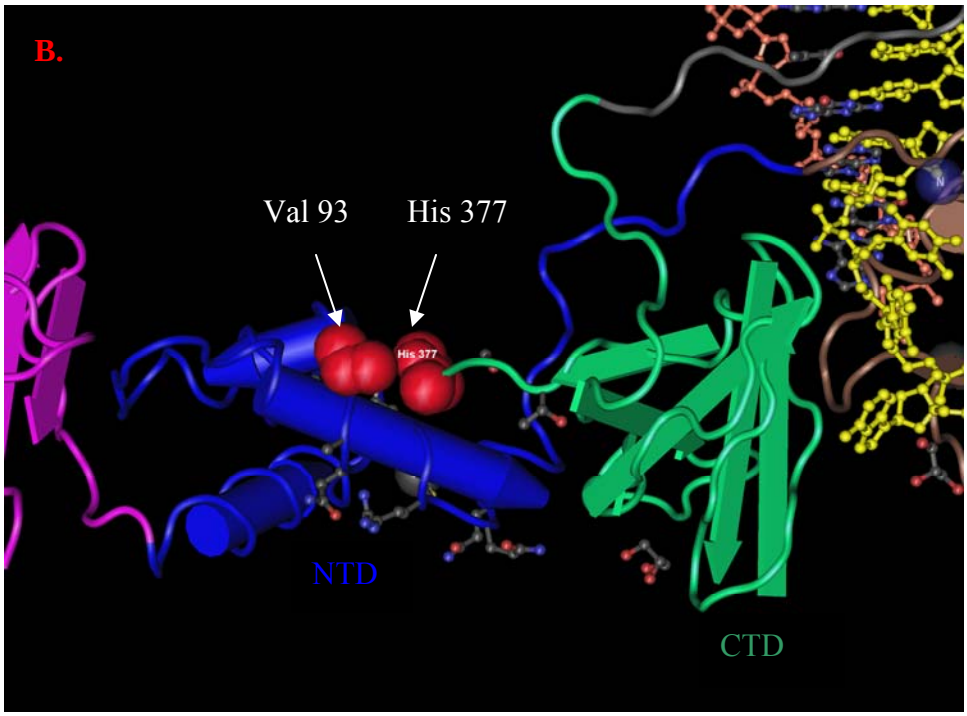
If the HIV-1 intasome does in fact assume a similar conformation to that of the PFV, then examination of the PFV crystal structure could reveal why certain truncations cannot be efficiently complemented for the strand transfer reaction. It is established that the IN subunit that performs 3' processing subsequently catalyzes the strand transfer reaction. Therefore in the observed complementation of the IN 266 and 267 with D116A for 3' processing, the truncated subunits are thus positioned as the internal subunits that will conduct the strand transfer reaction. The structure of the PFV intasome reveals that the CTD of one inner subunit is positioned between the CCD and NTD (Figure 44) where it interacts with both inner subunit CCDs and the phosphodiester backbone of both viral DNA molecules, thus serving to crosslink the two dimers. Intriguingly, the His 377 positioned at the end of the CTD SH3 fold comes into intimate contact with Val 93 in the NTD (Figure 44B). Also, the Ser 376 established Van der Waals interactions with the phosphodiester backbone of the unprocessed strand in the second DNA end [199], not depicted here. Thus one can envisage removal of these two residues from the extreme CTD terminal end could destabilize this region of the complex, leading to detrimental strand transfer.

In the past, analyses of IN truncations have been conducted, but to our knowledge never with the single amino acid resolution as presented here. A recent study of IN CTD truncations was reported, but the mutagenic survey was conducted at intervals of three to four residues, thus by-passing the IN 269 mutant for description [202]. Also our Class I

designation of the IN 270 truncation is in disagreement with their Class II assessment for this particular mutant. We do however observe and concur that the CTD tail is not critical for viral replication, apparently enhances both secondary and primary integrase functions with effectiveness concordant with its length. The functional boundaries within the CTD for specific IN activities and contributions to viral dynamics in the context of infection have been identified (Figure 45), and can be used as a stepping stone to further characterize IN primary and secondary roles. Studies of the IN 268/269 boundary for instance might provide insight into the poorly understood dynamics of precursor processing, and the mysterious role for IN in the context of within these processes. Also, by virtue of its complete lack of the unstructured tail, the enzymatically-active IN 269 mutant is the ideal tool for deciphering the contributions of the C-terminal tail to IN functionality. In this regard, uncovering the specific defect/s of IN 270, 271 and 272 will also provide a clearer picture of the role of the tail, and of the domain as a whole. These lines of inquiry will provide a more comprehensive knowledge of IN and its diverse roles in viral replication, which then lays a foundation for the potential development of multi-pronged therapeutic interventions that neutralize IN not only at integration, but also in its ancillary yet significant roles in other aspects of the viral life cycle.

Figure 44. Interactions of the PFV CTD within the intasome may give insight into HIV-1 CTD truncations. A**)** The CTD (green) of the PFV IN lies between the NTD (blue) and CCD (brown) in the tetrameric intasome, establishing contacts with both viral DNA ends, and contributing protein-protein interactions with the opposing subunit. One dimeric subunit is shown here. The monomer (gray) not participating in DNA contacts was only resolved for the CCD, the NTD and CTD remaining unresolved in the structure. The CTD participating in the reaction is colored green, and contacts both the viral DNA (yellow and pink ball and stick) and the NTD of its own subunit (blue). B) A close-up of the interaction between the very end of the CTD (His 377) with Val 93 in the NTD. The pertinent residues are shown in space-fill.


A.
Viral DNA
Monomer 1
Monomer 2


B.
Val 93
His 377
NTD
CTD

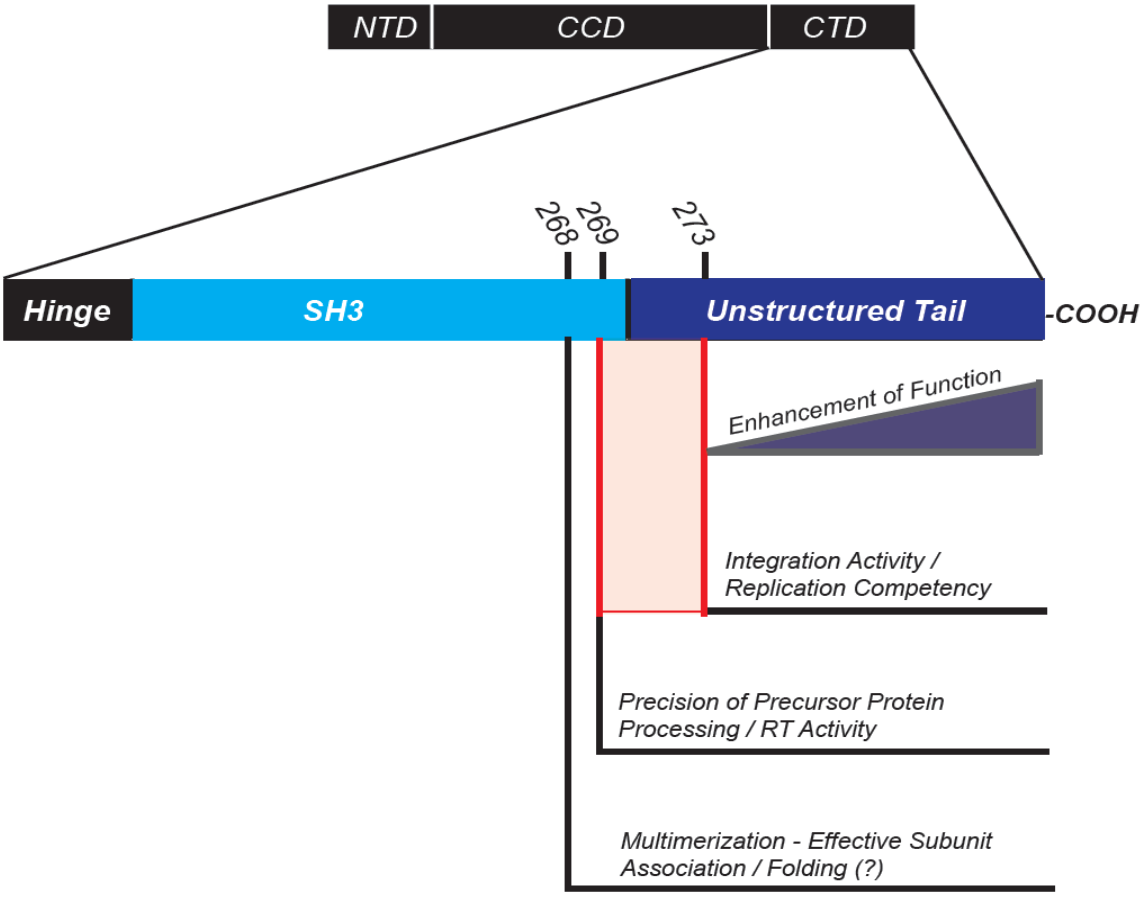


Figure 45. The properties of IN bestowed by increasing CTD length. A summary of the reported findings of the properties of IN at various CTD lengths.

CHAPTER 5- FUTURE STUDIES

ROLE OF LEDGF IN 3' PROCESSING

The association of HIV-1 IN with the LEDGF/p75 co-factor is one of the most well characterized interactions at the virus/host molecular interface. The predominantly nuclear LEDGF/p75 has been demonstrated to be essential during the early stage of infection, specifically at integration, asserting its effect by direct binding to IN. Despite the intensity of research, some ambiguity concerning the role of this protein in facilitating integration activity still exists. It has been suggested that its karyophilic properties may assist in PIC nuclear import, or that its action as a molecular tether targets IN to the chromosome by simultaneously binding chromosomal DNA and IN. *In vitro*, LEDGF enhances integration activity, and increases the solubility of recombinant IN. Several lines of evidence indicate that LEDGF binds a tetrameric form of IN, and so enhances the efficiency of concerted integration. LEDGF may thus also play a role in stabilizing IN complexes, functioning as a molecular chaperone to ensure the accurate assembly of IN oligomers required for integration, and by preventing the formation of higher-order non-productive aggregations of IN.

An additional element of uncertainty addresses the point at which LEDGF first encounters IN during infection, whether the interaction is first established in the nucleus, or in keeping with a role in nuclear import, within the cytoplasm. In one study, LEDGF was found as a component of cytoplasmic PICs in acutely infected cells [198] and thus the protein may in fact play a role in events occurring prior to or during nuclear import. The contribution of LEDGF to the 3' processing reaction is unclear, but its presence in

the PIC indicates that it could influence this first half of the integration reaction. The quantitative 3' processing assay we developed allows the opportunity to study the effect of LEDGF on viral DNA processing *in vivo*, by using cell-lines that stably express shRNAs that abolish LEDGF production. Though it is currently believed that a dimer of IN conducts the processing reaction at either end, a tetrameric model cannot be discounted. These experiments might thus illuminate both the role of LEDGF during HIV-1 infection, and the multimeric form of IN during the 3' processing reaction.

ORF VIRUS

The IN-VIF separated reading frame viruses have to this point only been preliminarily tested for replication competency. Though we found that these viruses replicated rather efficiently in primary cells, our manipulations have clearly disrupted the delicate balance of viral transcript splicing. The repercussions of this imbalance are unknown, but any effects, innocuous or not, detract from the validity of this virus as an investigative tool for deciphering the role of the IN CTD tail. We thus intend to serially-passage the virus through Vif non-permissive cells, to allow for the selection of fitness-enhanced viral variants that have presumptively fine-tuned the aberrant splicing program, plus any other defects we may have overlooked. The resulting virus then can be used for investigating the contribution of the tail in primary cell cultures of the natural targets of HIV-1, primary CD4+ T cells and macrophages, which require Vif function, and in which there may be a stricter and more discernable requirement for the IN tail.

The architectural manipulation of these viruses will permit investigation into the contribution of the lysine at position 273 in IN to viral dynamics, a residue found to be potently acetylated, and possibly subject to ubiquitination. The location of this residue in the ORF overlap between *pol(in)* and *vif* has hindered the investigation of these phenomena by mutational studies in primary cells due to the potential alteration of Vif protein function. The role of acetylation of IN is currently unclear, but given the divergent activities of the CTD in DNA binding and protein-protein interaction, it is intriguing that this domain is subject to a posttranslational modification known to modulate several aspects of protein function in a variety of systems.

DECIPHERING THE CONTRIBUTION OF THE C56Y AND I60T MUTATIONS

The second-site mutations C56Y and I60T enhance the efficiency of integration in the context of a tail-less IN, but the mechanism by which they do so has yet to be determined. Establishing the roles of these residues in the context of the integration complex would by extension illuminate the function of the tail during integration. To this end we will first ascertain the specific context of the compensatory mutations within the IN assembly that conducts the 3' processing reaction. Our model for the events occurring during processing (Figure 43) stipulates that the contribution of the second site mutations would only be relevant within the IN subunit that actually conducts the processing. We can verify this model by creating hybrid viruses containing a tail-less, catalytically defective IN D116A (IN D116A 269 stop) and an IN 269 stop IN with either of the second site mutations, and comparing these with hybrids that possess the second site

mutation in both IN molecules. Similar kinetics of processing in both cases should solidify that the basis for the rescued activity divulged by the compensatory mutations is within the specific context of the catalytically-active subunit. Additionally, several other combinatorial iterations of tail-less and full-length IN hybrid viruses can envisaged to further investigate the organization of IN within the processing and strand transfer complexes.

We postulated that the C56Y and I60T mutations might serve to rescue a DNA-binding defect that is wrought by tail truncation. We can test the DNA binding ability of the various truncation mutants, with and without the compensatory mutations, using *in vitro* systems to measure the affinity of recombinant IN molecules for viral DNA substrate. We will also test the catalytic activities of IN molecules containing these compensatory mutation using established *in vitro* assays for measuring 3' processing and strand transfer. Such experiments applied to the Class I IN mutants IN 270, 271 and 272, which we believe are defective for DNA binding, may identify the specific defects of these mutants.

PRECISION OF PARTICLE FORMATION

Precursor protein processing is initialized when Gag-Pol precursors dimerize, leading to activation of the embedded PR, and an ordered pattern of internal cleavage events within the precursor. The precursor processing behavior of the Class II IN truncations observed in Figure 27 may be attributable both to enhanced dimerization, and/or aberrant selection of initial PR cleavage sites that result in the premature and

detrimental accumulation of mature PR in the cytoplasm. It is well established that IN mutations and truncations can promote premature precursor processing, but the mechanism underlying the role of IN within this process is obscure. Our observations of the cleavage pattern of IN Class II mutants in the presence of high concentrations of Ritonavir suggest that the initial cleavages of the Gag-Pol precursor might be altered, and thus may account for the aberrant pattern of precursor protein products detected in these mutants. To study this phenomenon, we can determine the precise sites of initial cleavage of the viral precursor, by systematic substitution of a β-branched amino acid (such as isoleucine) at the P1 position of various PR cleavage sites. The P1 position is the first residue upstream of the scissile bond, and mutation of this position has been shown to abolish PR cleavage [203]. The resulting discernment of an altered pattern of initial cleavages afforded by incubation in high concentrations of Ritonavir will permit the identification of the alternate sites that are utilized by the class II versus WT IN truncations.

CTD INTERACTIONS WITH KNOWN BINDING PARTNERS

Several binding partners have been found to interact with IN via the CTD. These include EED polycomb, p300, HIV-1 RT, Rad18 [3, 110, 146-148]. The ability of the different truncations, especially in the IN 268 269 271 272 273 region will be tested for differential affinities for interacting with these partners, via reciprocal pull-down experiments or by microscopic techniques. The loss or gain of interaction may reveal the significance if any of any of these controversial IN interactors.

REFERENCES

1. Bokazhanova, A. and G.W. Rutherford, *The epidemiology of HIV and AIDS in the world.* Coll Antropol, 2006. **30 Suppl 2**: p. 3-10.

2. Lu, R., H.Z. Ghory, and A. Engelman, *Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase.* J Virol, 2005. **79**(16): p. 10356-68.

3. Cereseto, A., et al., *Acetylation of HIV-1 integrase by p300 regulates viral integration.* EMBO J, 2005. **24**(17): p. 3070-81.

4. Topper, M., et al., *Posttranslational acetylation of the human immunodeficiency virus type 1 integrase carboxyl-terminal domain is dispensable for viral replication.* J Virol, 2007. **81**(6): p. 3012-7.

5. Fischle, W., Y. Wang, and C.D. Allis, *Histone and chromatin cross-talk.* Curr Opin Cell Biol, 2003. **15**(2): p. 172-83.

6. Gonda, M.A., et al., *Human T-cell lymphotropic virus type III shares sequence homology with a family of pathogenic lentiviruses.* Proc Natl Acad Sci U S A, 1986. **83**(11): p. 4007-11.

7. Gonda, M.A., et al., *Sequence homology and morphologic similarity of HTLV-III and visna virus, a pathogenic lentivirus.* Science, 1985. **227**(4683): p. 173-7.

8. Pollard, V.W. and M.H. Malim, *The HIV-1 Rev protein.* Annu Rev Microbiol, 1998. **52**: p. 491-532.

9. Perelson, A.S., et al., *HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time.* Science, 1996. **271**(5255): p. 1582-6.

10. Tsai, W.P., et al., *Preliminary in vitro growth cycle and transmission studies of HIV-1 in an autologous primary cell assay of blood-derived macrophages and peripheral blood mononuclear cells.* Virology, 1996. **226**(2): p. 205-16.

11. Eckstein, D.A., et al., *HIV-1 actively replicates in naive CD4(+) T cells residing within human lymphoid tissues.* Immunity, 2001. **15**(4): p. 671-82.

12. Dimitrov, D.S., et al., *Quantitation of human immunodeficiency virus type 1 infection kinetics.* J Virol, 1993. **67**(4): p. 2182-90.

13. Chen, H.Y., et al., *Determination of virus burst size in vivo using a single-cycle SIV in rhesus macaques.* Proc Natl Acad Sci U S A, 2007. **104**(48): p. 19079-84.

14. Michael, N.L. and J.P. Moore, *HIV-1 entry inhibitors: evading the issue.* Nat Med, 1999. **5**(7): p. 740-2.

15. Kao, S.Y., et al., *Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product.* Nature, 1987. **330**(6147): p. 489-93.

16. Efthymiadis, A., L.J. Briggs, and D.A. Jans, *The HIV-1 Tat nuclear localization sequence confers novel nuclear import properties.* J Biol Chem, 1998. **273**(3): p. 1623-8.

17. Truant, R. and B.R. Cullen, *The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals.* Mol Cell Biol, 1999. **19**(2): p. 1210-7.

18. Muesing, M.A., D.H. Smith, and D.J. Capon, *Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein.* Cell, 1987. **48**(4): p. 691-701.

19. Berkhout, B. and K.T. Jeang, *trans activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the trans-acting-responsive hairpin: a quantitative analysis.* J Virol, 1989. **63**(12): p. 5501-4.

20. Berkhout, B., R.H. Silverman, and K.T. Jeang, *Tat trans-activates the human immunodeficiency virus through a nascent RNA target.* Cell, 1989. **59**(2): p. 273-82.

21. Brady, J. and F. Kashanchi, *Tat gets the "green" light on transcription initiation.* Retrovirology, 2005. **2**: p. 69.

22. Legrain, P. and M. Rosbash, *Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm.* Cell, 1989. **57**(4): p. 573-83.

23. Chang, D.D. and P.A. Sharp, *Regulation by HIV Rev depends upon recognition of splice sites.* Cell, 1989. **59**(5): p. 789-95.

24. Nakielny, S., et al., *RNA transport.* Annu Rev Neurosci, 1997. **20**: p. 269-301.

25. Daly, T.J., et al., *Biochemical characterization of binding of multiple HIV-1 Rev monomeric proteins to the Rev responsive element.* Biochemistry, 1993. **32**(39): p. 10497-505.

26. Daly, T.J., et al., *Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro.* Nature, 1989. **342**(6251): p. 816-9.

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

28. Haffar, O., et al., *Human immunodeficiency virus-like, nonreplicating, gag-env particles assemble in a recombinant vaccinia virus expression system.* J Virol, 1990. **64**(6): p. 2653-9.

29. Karacostas, V., et al., *Human immunodeficiency virus-like particles produced by a vaccinia virus expression vector.* Proc Natl Acad Sci U S A, 1989. **86**(22): p. 8964-7.

30. Smith, A.J., et al., *Human immunodeficiency virus type 1 Pr55gag and Pr160gag-pol expressed from a simian virus 40 late replacement vector are efficiently processed and assembled into viruslike particles.* J Virol, 1990. **64**(6): p. 2743-50.

31. Bryant, M. and L. Ratner, *Myristoylation-dependent replication and assembly of human immunodeficiency virus 1.* Proc Natl Acad Sci U S A, 1990. **87**(2): p. 523-7.

32. Garrus, J.E., et al., *Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding.* Cell, 2001. **107**(1): p. 55-65.

33. VerPlank, L., et al., *Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag).* Proc Natl Acad Sci U S A, 2001. **98**(14): p. 7724-9.

34. Jacks, T., et al., *Characterization of ribosomal frameshifting in HIV-1 gag-pol expression.* Nature, 1988. **331**(6153): p. 280-3.

35. Brierley, I. and F.J. Dos Ramos, *Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV.* Virus Res, 2006. **119**(1): p. 29-42.

36. Karacostas, V., et al., *Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles.* Virology, 1993. **193**(2): p. 661-71.

37. Huang, M. and M.A. Martin, *Incorporation of Pr160(gag-pol) into virus particles requires the presence of both the major homology region and adjacent C-terminal capsid sequences within the Gag-Pol polyprotein.* J Virol, 1997. **71**(6): p. 4472-8.

38. Srinivasakumar, N., M.L. Hammarskjold, and D. Rekosh, *Characterization of deletion mutations in the capsid region of human immunodeficiency virus type 1 that affect particle formation and Gag-Pol precursor incorporation.* J Virol, 1995. **69**(10): p. 6106-14.

39. Stein, B.S. and E.G. Engleman, *Intracellular processing of the gp160 HIV-1 envelope precursor. Endoproteolytic cleavage occurs in a cis or medial compartment of the Golgi complex.* J Biol Chem, 1990. **265**(5): p. 2640-9.

40. Gu, M., J. Rappaport, and S.H. Leppla, *Furin is important but not essential for the proteolytic maturation of gp160 of HIV-1.* FEBS Lett, 1995. **365**(1): p. 95-7.

41. Dannull, J., et al., *Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger and flanking basic amino acid residues.* EMBO J, 1994. **13**(7): p. 1525-33.

42. Aldovini, A. and R.A. Young, *Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus.* J Virol, 1990. **64**(5): p. 1920-6.

43. Lever, A., et al., *Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions.* J Virol, 1989. **63**(9): p. 4085-7.

44. Neil, S.J., et al., *An interferon-alpha-induced tethering mechanism inhibits HIV-1 and Ebola virus particle release but is counteracted by the HIV-1 Vpu protein.* Cell Host Microbe, 2007. **2**(3): p. 193-203.

45. Neil, S.J., T. Zang, and P.D. Bieniasz, *Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu.* Nature, 2008. **451**(7177): p. 425-30.

46. Van Damme, N., et al., *The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein.* Cell Host Microbe, 2008. **3**(4): p. 245-52.

47. Bukovsky, A. and H. Gottlinger, *Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease.* J Virol, 1996. **70**(10): p. 6820-5.

48. Quillent, C., et al., *Extensive regions of pol are required for efficient human immunodeficiency virus polyprotein processing and particle maturation.* Virology, 1996. **219**(1): p. 29-36.

49. Pettit, S.C., et al., *Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism.* J Virol, 2004. **78**(16): p. 8477-85.

50. Dorfman, T., et al., *Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein.* J Virol, 1994. **68**(3): p. 1689-96.

51. Alkhatib, G., et al., *CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1.* Science, 1996. **272**(5270): p. 1955-8.

52. Choe, H., et al., *The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates.* Cell, 1996. **85**(7): p. 1135-48.

53. Doranz, B.J., et al., *A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors.* Cell, 1996. **85**(7): p. 1149-58.

54. Dragic, T., et al., *HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5.* Nature, 1996. **381**(6584): p. 667-73.

55. Feng, Y., et al., *HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor.* Science, 1996. **272**(5263): p. 872-7.

56. Combadiere, C., et al., *Cloning and functional expression of CC CKR5, a human monocyte CC chemokine receptor selective for MIP-1(alpha), MIP-1(beta), and RANTES.* J Leukoc Biol, 1996. **60**(1): p. 147-52.

57. Hwang, S.S., et al., *Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1.* Science, 1991. **253**(5015): p. 71-4.

58. Fassati, A. and S.P. Goff, *Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1.* J Virol, 2001. **75**(8): p. 3626-35.

59. Mulky, A., et al., *Subunit-specific analysis of the human immunodeficiency virus type 1 reverse transcriptase in vivo.* J Virol, 2004. **78**(13): p. 7089-96.

60. Jiang, M., et al., *Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1.* J Virol, 1993. **67**(6): p. 3246-53.

61. Mak, J. and L. Kleiman, *Primer tRNAs for reverse transcription.* J Virol, 1997. **71**(11): p. 8087-95.

62. Mansky, L.M., *The mutation rate of human immunodeficiency virus type 1 is influenced by the vpr gene.* Virology, 1996. **222**(2): p. 391-400.

63. Mansky, L.M. and H.M. Temin, *Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase.* J Virol, 1995. **69**(8): p. 5087-94.

64. Lewis, P., M. Hensel, and M. Emerman, *Human immunodeficiency virus infection of cells arrested in the cell cycle.* EMBO J, 1992. **11**(8): p. 3053-8.

65. Valentin, A., et al., *In vitro maturation of mononuclear phagocytes and susceptibility to HIV-1 infection.* J Acquir Immune Defic Syndr, 1991. **4**(8): p. 751-9.

66. Ho, D.D., T.R. Rota, and M.S. Hirsch, *Infection of monocyte/macrophages by human T lymphotropic virus type III.* J Clin Invest, 1986. **77**(5): p. 1712-5.

67. Lewis, P.F. and M. Emerman, *Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus.* J Virol, 1994. **68**(1): p. 510-6.

68. Roe, T., et al., *Integration of murine leukemia virus DNA depends on mitosis.* EMBO J, 1993. **12**(5): p. 2099-108.

69. Humphries, E.H. and H.M. Temin, *Requirement for cell division for initiation of transcription of Rous sarcoma virus RNA.* J Virol, 1974. **14**(3): p. 531-46.

70. Bukrinsky, M.I., et al., *Active nuclear import of human immunodeficiency virus type 1 preintegration complexes.* Proc Natl Acad Sci U S A, 1992. **89**(14): p. 6580-4.

71. Heinzinger, N.K., et al., *The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells.* Proc Natl Acad Sci U S A, 1994. **91**(15): p. 7311-5.

72. Gallay, P., et al., *HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway.* Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9825-30.

73. Bukrinsky, M.I., et al., *A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells.* Nature, 1993. **365**(6447): p. 666-9.

74. Zennou, V., et al., *HIV-1 genome nuclear import is mediated by a central DNA flap.* Cell, 2000. **101**(2): p. 173-85.

75. Pauza, C.D., J.E. Galindo, and D.D. Richman, *Reinfection results in accumulation of unintegrated viral DNA in cytopathic and persistent human immunodeficiency virus type 1 infection of CEM cells.* J Exp Med, 1990. **172**(4): p. 1035-42.

76. Pang, S., et al., *High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients.* Nature, 1990. **343**(6253): p. 85-9.

77. Teo, I., et al., *Circular forms of unintegrated human immunodeficiency virus type 1 DNA and high levels of viral protein expression: association with dementia and multinucleated giant cells in the brains of patients with AIDS.* J Virol, 1997. **71**(4): p. 2928-33.

78. Dina, D. and E.W. Benz, Jr., *Structure of murine sarcoma virus DNA replicative intermediates synthesized in vitro.* J Virol, 1980. **33**(1): p. 377-89.

79. Junghans, R.P., L.R. Boone, and A.M. Skalka, *Products of reverse transcription in avian retrovirus analyzed by electron microscopy.* J Virol, 1982. **43**(2): p. 544-54.

80. Gilboa, E., et al., *In vitro synthesis of a 9 kbp terminally redundant DNA carrying the infectivity of Moloney murine leukemia virus.* Cell, 1979. **16**(4): p. 863-74.

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

82. Li, L., et al., *Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection.* EMBO J, 2001. **20**(12): p. 3272-81.

83. Kelly, J., et al., *Human macrophages support persistent transcription from unintegrated HIV-1 DNA.* Virology, 2008. **372**(2): p. 300-12.

84. Malim, M.H. and M. Emerman, *HIV-1 accessory proteins--ensuring viral survival in a hostile environment.* Cell Host Microbe, 2008. **3**(6): p. 388-98.

85. Madrid, R., et al., *Nef-induced alteration of the early/recycling endosomal compartment correlates with enhancement of HIV-1 infectivity.* J Biol Chem, 2005. **280**(6): p. 5032-44.

86. Greenway, A.L., et al., *Human immunodeficiency virus type 1 Nef binds to tumor suppressor p53 and protects cells against p53-mediated apoptosis.* J Virol, 2002. **76**(6): p. 2692-702.

87. Xu, X.N., et al., *Induction of Fas ligand expression by HIV involves the interaction of Nef with the T cell receptor zeta chain.* J Exp Med, 1999. **189**(9): p. 1489-96.

88. Lama, J., A. Mangasarian, and D. Trono, *Cell-surface expression of CD4 reduces HIV-1 infectivity by blocking Env incorporation in a Nef- and Vpu-inhibitable manner.* Curr Biol, 1999. **9**(12): p. 622-31.

89. Schwartz, O., et al., *Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein.* Nat Med, 1996. **2**(3): p. 338-42.

90. Ross, T.M., A.E. Oran, and B.R. Cullen, *Inhibition of HIV-1 progeny virion release by cell-surface CD4 is relieved by expression of the viral Nef protein.* Curr Biol, 1999. **9**(12): p. 613-21.

91. Malim, M.H., *APOBEC proteins and intrinsic resistance to HIV-1 infection.* Philos Trans R Soc Lond B Biol Sci, 2009. **364**(1517): p. 675-87.

92. Goila-Gaur, R. and K. Strebel, *HIV-1 Vif, APOBEC, and intrinsic immunity.* Retrovirology, 2008. **5**: p. 51.

93. Vodicka, M.A., et al., *HIV-1 Vpr interacts with the nuclear transport pathway to promote macrophage infection.* Genes Dev, 1998. **12**(2): p. 175-85.

94. Waldhuber, M.G., et al., *Studies with GFP-Vpr fusion proteins: induction of apoptosis but ablation of cell-cycle arrest despite nuclear membrane or nuclear localization.* Virology, 2003. **313**(1): p. 91-104.

95. Bachand, F., et al., *Incorporation of Vpr into human immunodeficiency virus type 1 requires a direct interaction with the p6 domain of the p55 gag precursor.* J Biol Chem, 1999. **274**(13): p. 9083-91.

96. Mansky, L.M., et al., *The interaction of vpr with uracil DNA glycosylase modulates the human immunodeficiency virus type 1 In vivo mutation rate.* J Virol, 2000. **74**(15): p. 7039-47.

97. Yasuda, J., et al., *T cell apoptosis causes peripheral T cell depletion in mice transgenic for the HIV-1 vpr gene.* Virology, 2001. **285**(2): p. 181-92.

98. Ayyavoo, V., et al., *HIV-1 Vpr suppresses immune activation and apoptosis through regulation of nuclear factor kappa B.* Nat Med, 1997. **3**(10): p. 1117-23.

99. Gummuluru, S. and M. Emerman, *Cell cycle- and Vpr-mediated regulation of human immunodeficiency virus type 1 expression in primary and transformed T-cell lines.* J Virol, 1999. **73**(7): p. 5422-30.

100. Poon, B., M.A. Chang, and I.S. Chen, *Vpr is required for efficient Nef expression from unintegrated human immunodeficiency virus type 1 DNA.* J Virol, 2007. **81**(19): p. 10515-23.

101. Poon, B. and I.S. Chen, *Human immunodeficiency virus type 1 (HIV-1) Vpr enhances expression from unintegrated HIV-1 DNA.* J Virol, 2003. **77**(7): p. 3962-72.

102. Jowett, J.B., et al., *The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle.* J Virol, 1995. **69**(10): p. 6304-13.

103. Eckstein, D.A., et al., *HIV-1 Vpr enhances viral burden by facilitating infection of tissue macrophages but not nondividing CD4+ T cells.* J Exp Med, 2001. **194**(10): p. 1407-19.

104. Goh, W.C., et al., *HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo.* Nat Med, 1998. **4**(1): p. 65-71.

105. Yao, X.J., et al., *The effect of vpu on HIV-1-induced syncytia formation.* J Acquir Immune Defic Syndr, 1993. **6**(2): p. 135-41.

106. Wiskerchen, M. and M.A. Muesing, *Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells.* J Virol, 1995. **69**(1): p. 376-86.

107. Engelman, A., et al., *Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication.* J Virol, 1995. **69**(5): p. 2729-36.

108. Dobard, C.W., M.S. Briones, and S.A. Chow, *Molecular mechanisms by which human immunodeficiency virus type 1 integrase stimulates the early steps of reverse transcription.* J Virol, 2007. **81**(18): p. 10037-46.

109. Wu, X., et al., *Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions with the nucleoprotein reverse transcription complex.* J Virol, 1999. **73**(3): p. 2126-35.

110. Zhu, K., C. Dobard, and S.A. Chow, *Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase.* J Virol, 2004. **78**(10): p. 5045-55.

111. Leavitt, A.D., et al., *Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection.* J Virol, 1996. **70**(2): p. 721-8.

112. Masuda, T., et al., *Genetic analysis of human immunodeficiency virus type 1 integrase and the U3 att site: unusual phenotype of mutants in the zinc finger-like domain.* J Virol, 1995. **69**(11): p. 6687-96.

113. Tsurutani, N., et al., *Identification of critical amino acid residues in human immunodeficiency virus type 1 IN required for efficient proviral DNA formation at steps prior to integration in dividing and nondividing cells.* J Virol, 2000. **74**(10): p. 4795-806.

114. Engelman, A., *In vivo analysis of retroviral integrase structure and function.* Adv Virus Res, 1999. **52**: p. 411-26.

115. Savarino, A., *In-Silico docking of HIV-1 integrase inhibitors reveals a novel drug type acting on an enzyme/DNA reaction intermediate.* Retrovirology, 2007. **4**: p. 21.

116. Dyda, F., et al., *Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases.* Science, 1994. **266**(5193): p. 1981-6.

117. Yang, W. and T.A. Steitz, *Recombining the structures of HIV integrase, RuvC and RNase H.* Structure, 1995. **3**(2): p. 131-4.

118. Rice, P., R. Craigie, and D.R. Davies, *Retroviral integrases and their cousins.* Curr Opin Struct Biol, 1996. **6**(1): p. 76-83.

119. Bujacz, G., et al., *The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant.* FEBS Lett, 1996. **398**(2-3): p. 175-8.

120. Engelman, A. and R. Craigie, *Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro.* J Virol, 1992. **66**(11): p. 6361-9.

121. Engelman, A., F.D. Bushman, and R. Craigie, *Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex.* EMBO J, 1993. **12**(8): p. 3269-75.

122. van Gent, D.C., et al., *Complementation between HIV integrase proteins mutated in different domains.* EMBO J, 1993. **12**(8): p. 3261-7.

123. Cai, M., et al., *Solution structure of the N-terminal zinc binding domain of HIV-1 integrase.* Nat Struct Biol, 1997. **4**(7): p. 567-77.

124. Chen, J.C., et al., *Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding.* Proc Natl Acad Sci U S A, 2000. **97**(15): p. 8233-8.

125. Eijkelenboom, A.P., et al., *Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase.* Proteins, 1999. **36**(4): p. 556-64.

126. Eijkelenboom, A.P., et al., *The DNA-binding domain of HIV-1 integrase has an SH3-like fold.* Nat Struct Biol, 1995. **2**(9): p. 807-10.

127. Goldgur, Y., et al., *Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9150-4.

128. Lodi, P.J., et al., *Solution structure of the DNA binding domain of HIV-1 integrase.* Biochemistry, 1995. **34**(31): p. 9826-33.

129. Wang, J.Y., et al., *Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein.* EMBO J, 2001. **20**(24): p. 7333-43.

130. Burke, C.J., et al., *Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase.* J Biol Chem, 1992. **267**(14): p. 9639-44.

131. Zheng, R., T.M. Jenkins, and R. Craigie, *Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13659-64.

132. Zhao, Z., et al., *Subunit-specific protein footprinting reveals significant structural rearrangements and a role for N-terminal Lys-14 of HIV-1 Integrase during viral DNA binding.* J Biol Chem, 2008. **283**(9): p. 5632-41.

133. Rowland, S.J. and K.G. Dyke, *Tn552, a novel transposable element from Staphylococcus aureus.* Mol Microbiol, 1990. **4**(6): p. 961-75.

134. Polard, P. and M. Chandler, *Bacterial transposases and retroviral integrases.* Mol Microbiol, 1995. **15**(1): p. 13-23.

135. Esposito, D. and R. Craigie, *Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction.* EMBO J, 1998. **17**(19): p. 5832-43.

136. Kulkosky, J., et al., *Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases.* Mol Cell Biol, 1992. **12**(5): p. 2331-8.

137. Drelich, M., R. Wilhelm, and J. Mous, *Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein in vitro.* Virology, 1992. **188**(2): p. 459-68.

138. Craigie, R., *HIV integrase, a brief overview from chemistry to therapeutics.* J Biol Chem, 2001. **276**(26): p. 23213-6.

139. Lutzke, R.A., C. Vink, and R.H. Plasterk, *Characterization of the minimal DNA-binding domain of the HIV integrase protein.* Nucleic Acids Res, 1994. **22**(20): p. 4125-31.

140. Emiliani, S., et al., *Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication.* J Biol Chem, 2005. **280**(27): p. 25517-23.

141. Busschots, K., et al., *The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding.* J Biol Chem, 2005. **280**(18): p. 17841-7.

142. Jenkins, T.M., et al., *A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization.* J Biol Chem, 1996. **271**(13): p. 7712-8.

143. Ao, Z., et al., *Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication.* J Biol Chem, 2007. **282**(18): p. 13456-67.

144. Hamamoto, S., et al., *Identification of a novel human immunodeficiency virus type 1 integrase interactor, Gemin2, that facilitates efficient viral cDNA synthesis in vivo.* J Virol, 2006. **80**(12): p. 5670-7.

145. Luo, K., et al., *Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation.* J Virol, 2007. **81**(13): p. 7238-48.

146. Mulder, L.C., L.A. Chakrabarti, and M.A. Muesing, *Interaction of HIV-1 integrase with DNA repair protein hRad18.* J Biol Chem, 2002. **277**(30): p. 27489-93.

147. Violot, S., et al., *The human polycomb group EED protein interacts with the integrase of human immunodeficiency virus type 1.* J Virol, 2003. **77**(23): p. 12507-22.

148. Wilkinson, T.A., et al., *Identifying and characterizing a functional HIV-1 reverse transcriptase-binding site on integrase.* J Biol Chem, 2009. **284**(12): p. 7931-9.

149. Hehl, E.A., et al., *Interaction between human immunodeficiency virus type 1 reverse transcriptase and integrase proteins.* J Virol, 2004. **78**(10): p. 5056-67.

150. Schrofelbauer, B., et al., *Mutational alteration of human immunodeficiency virus type 1 Vif allows for functional interaction with nonhuman primate APOBEC3G.* J Virol, 2006. **80**(12): p. 5984-91.

151. Tian, C., et al., *Differential requirement for conserved tryptophans in human immunodeficiency virus type 1 Vif for the selective suppression of APOBEC3G and APOBEC3F.* J Virol, 2006. **80**(6): p. 3112-5.

152. Wichroski, M.J., K. Ichiyama, and T.M. Rana, *Analysis of HIV-1 viral infectivity factor-mediated proteasome-dependent depletion of APOBEC3G: correlating function and subcellular localization.* J Biol Chem, 2005. **280**(9): p. 8387-96.

153. Mousnier, A., et al., *von Hippel Lindau binding protein 1-mediated degradation of integrase affects HIV-1 gene expression at a postintegration step.* Proc Natl Acad Sci U S A, 2007. **104**(34): p. 13615-20.

154. Sinha, S., M.H. Pursley, and D.P. Grandgenett, *Efficient concerted integration by recombinant human immunodeficiency virus type 1 integrase without cellular or viral cofactors.* J Virol, 2002. **76**(7): p. 3105-13.

155. Delelis, O., et al., *Integrase and integration: biochemical activities of HIV-1 integrase.* Retrovirology, 2008. **5**: p. 114.

156. Panganiban, A.T. and H.M. Temin, *The terminal nucleotides of retrovirus DNA are required for integration but not virus production.* Nature, 1983. **306**(5939): p. 155-60.

157. Faure, A., et al., *HIV-1 integrase crosslinked oligomers are active in vitro.* Nucleic Acids Res, 2005. **33**(3): p. 977-86.

158. Hayouka, Z., et al., *Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium.* Proc Natl Acad Sci U S A, 2007. **104**(20): p. 8316-21.

159. Guiot, E., et al., *Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity.* J Biol Chem, 2006. **281**(32): p. 22707-19.

160. Deprez, E., et al., *Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy.* Biochemistry, 2000. **39**(31): p. 9275-84.

161. Lutzke, R.A. and R.H. Plasterk, *Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding.* J Virol, 1998. **72**(6): p. 4841-8.

162. Diamond, T.L. and F.D. Bushman, *Division of labor within human immunodeficiency virus integrase complexes: determinants of catalysis and target DNA capture.* J Virol, 2005. **79**(24): p. 15376-87.

163. Gao, K., S.L. Butler, and F. Bushman, *Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes.* EMBO J, 2001. **20**(13): p. 3565-76.

164. Heuer, T.S. and P.O. Brown, *Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex.* Biochemistry, 1998. **37**(19): p. 6667-78.

165. Heuer, T.S. and P.O. Brown, *Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking.* Biochemistry, 1997. **36**(35): p. 10655-65.

166. Michel, F., et al., *Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor.* EMBO J, 2009. **28**(7): p. 980-91.

167. Cherepanov, P., et al., *HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells.* J Biol Chem, 2003. **278**(1): p. 372-81.

168. Pandey, K.K., S. Sinha, and D.P. Grandgenett, *Transcriptional coactivator LEDGF/p75 modulates human immunodeficiency virus type 1 integrase-mediated concerted integration.* J Virol, 2007. **81**(8): p. 3969-79.

169. Bushman, F.D., et al., *Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding.* Proc Natl Acad Sci U S A, 1993. **90**(8): p. 3428-32.

170. Park, J. and C.D. Morrow, *Overexpression of the gag-pol precursor from human immunodeficiency virus type 1 proviral genomes results in efficient proteolytic processing in the absence of virion production.* J Virol, 1991. **65**(9): p. 5111-7.

171. Tachedjian, G., et al., *Efavirenz enhances the proteolytic processing of an HIV-1 pol polyprotein precursor and reverse transcriptase homodimer formation.* FEBS Lett, 2005. **579**(2): p. 379-84.

172. Fletcher, T.M., 3rd, et al., *Complementation of integrase function in HIV-1 virions.* EMBO J, 1997. **16**(16): p. 5123-38.

173. Willetts, K.E., et al., *DNA repair enzyme uracil DNA glycosylase is specifically incorporated into human immunodeficiency virus type 1 viral particles through a Vpr-independent mechanism.* J Virol, 1999. **73**(2): p. 1682-8.

174. Yung, E., et al., *Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1.* Nat Med, 2001. **7**(8): p. 920-6.

175. Chen, R., et al., *Vpr-mediated incorporation of UNG2 into HIV-1 particles is required to modulate the virus mutation rate and for replication in macrophages.* J Biol Chem, 2004. **279**(27): p. 28419-25.

176. Miller, R.J., et al., *Human immunodeficiency virus and AIDS: insights from animal lentiviruses.* J Virol, 2000. **74**(16): p. 7187-95.

177. Wang, W., et al., *Purification and biochemical heterogeneity of the mammalian SWI-SNF complex.* EMBO J, 1996. **15**(19): p. 5370-82.

178. Sorin, M., et al., *Recruitment of a SAP18-HDAC1 complex into HIV-1 virions and its requirement for viral replication.* PLoS Pathog, 2009. **5**(6): p. e1000463.

179. Low, A., et al., *Natural polymorphisms of human immunodeficiency virus type 1 integrase and inherent susceptibilities to a panel of integrase inhibitors.* Antimicrob Agents Chemother, 2009. **53**(10): p. 4275-82.

180. Morner, A., et al., *Primary human immunodeficiency virus type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in coreceptor usage.* J Virol, 1999. **73**(3): p. 2343-9.

181. Charneau, P., et al., *HIV-1 reverse transcription. A termination step at the center of the genome.* J Mol Biol, 1994. **241**(5): p. 651-62.

182. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal Biochem, 1976. **72**: p. 248-54.

183. Nilsen, B.M., et al., *Monoclonal antibodies against human immunodeficiency virus type 1 integrase: epitope mapping and differential effects on integrase activities in vitro.* J Virol, 1996. **70**(3): p. 1580-7.

184. Chesebro, B., et al., *Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism.* J Virol, 1992. **66**(11): p. 6547-54.

185. Szilvay, A.M., et al., *Epitope mapping of HIV-1 reverse transcriptase with monoclonal antibodies that inhibit polymerase and RNase H activities.* J Acquir Immune Defic Syndr, 1992. **5**(7): p. 647-57.

186. Chun, T.W., et al., *Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy.* Proc Natl Acad Sci U S A, 1997. **94**(24): p. 13193-7.

187. Mueller, P.R. and B. Wold, *In vivo footprinting of a muscle specific enhancer by ligation mediated PCR.* Science, 1989. **246**(4931): p. 780-6.

188. Butler, S.L., M.S. Hansen, and F.D. Bushman, *A quantitative assay for HIV DNA integration in vivo.* Nat Med, 2001. **7**(5): p. 631-4.

189. Pettit, S.C., et al., *Ordered processing of the human immunodeficiency virus type 1 GagPol precursor is influenced by the context of the embedded viral protease.* J Virol, 2005. **79**(16): p. 10601-7.

190. Pettit, S.C., et al., *Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates.* Retrovirology, 2005. **2**: p. 66.

191. Li, F., et al., *PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13555-60.

192. Cen, S., et al., *Incorporation of pol into human immunodeficiency virus type 1 Gag virus-like particles occurs independently of the upstream Gag domain in Gag-pol.* J Virol, 2004. **78**(2): p. 1042-9.

193. Chiu, H.C., S.Y. Yao, and C.T. Wang, *Coding sequences upstream of the human immunodeficiency virus type 1 reverse transcriptase domain in Gag-Pol are not essential for incorporation of the Pr160(gag-pol) into virus particles.* J Virol, 2002. **76**(7): p. 3221-31.

194. Liao, W.H., et al., *Incorporation of human immunodeficiency virus type 1 reverse transcriptase into virus-like particles.* J Virol, 2007. **81**(10): p. 5155-65.

195. Mulder, L.C. and M.A. Muesing, *Degradation of HIV-1 integrase by the N-end rule pathway.* J Biol Chem, 2000. **275**(38): p. 29749-53.

196. Mogk, A., R. Schmidt, and B. Bukau, *The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies.* Trends Cell Biol, 2007. **17**(4): p. 165-72.

197. Llano, M., et al., *An essential role for LEDGF/p75 in HIV integration.* Science, 2006. **314**(5798): p. 461-4.

198. Llano, M., et al., *LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes.* J Virol, 2004. **78**(17): p. 9524-37.

199. Hare, S., et al., *Retroviral intasome assembly and inhibition of DNA strand transfer.* Nature. **464**(7286): p. 232-6.

200. Scottoline, B.P., et al., *Disruption of the terminal base pairs of retroviral DNA during integration.* Genes Dev, 1997. **11**(3): p. 371-82.

201. Balakrishnan, M. and C.B. Jonsson, *Functional identification of nucleotides conferring substrate specificity to retroviral integrase reactions.* J Virol, 1997. **71**(2): p. 1025-35.

202. Dar, M.J., et al., *Biochemical and virological analysis of the 18-residue C-terminal tail of HIV-1 integrase.* Retrovirology, 2009. **6**: p. 94.

203. Pettit, S.C., et al., *Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease.* J Virol, 2002. **76**(20): p. 10226-33.