2007

A Genetic-Proteomic Approach to Identify Cellular Components that Interact with HIV-1

Yang Luo

A GENETIC-PROTEOMIC APPROACH TO IDENTIFY CELLULAR COMPONENTS THAT INTERACT WITH HIV-1

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Yang Luo

June 2007

A Genetic-Proteomic Approach to Identify Cellular Components that Interact with HIV-1

Yang Luo, Ph.D.

The Rockefeller University, 2007

Given the limited genetic coding capacity of HIV-1, it is reasonable to expect that the virus must interact with an extensive set of cellular factors and their complexes to complete its passage through the cell. Indeed, it is remarkable that the viral genome, comprising only about 0.0003% of the entire genetic capacity of the cell, commandeers the cellular environment to its own advantage. However, to date, only a small group of cellular proteins have been shown to be required for viral propagation. In an effort to recover and identify those host proteins that interact in complex with the viral machinery, we have developed a systematic genetic method to select derivatives that can encode a small, but potent, foreign epitope tag yet remain fully replication-competent in culture. In conjunction with a novel cryogenic methodology to capture and preserve viral-host interactions usually lost when more conventional isolation techniques are employed, we have recovered host complexes that interact specifically with each of three independently tagged HIV-1 proteins during progressive infection. Thus, the quantitative purification of the tagged viral proteins has allowed the identification of both described factors already known to interact with each of the

targeted viral proteins and as well, unanticipated sets of new host proteins in complex with the virus and previously obscured from investigation. Identification and characterization of protein-protein interactions between the host and the virus will provide insight into the cellular processes expropriated by the virus to complete its life cycle.

献给我的父亲母亲

In memory of my father, Yinsheng Luo

To the love of my mother, Huazhen Liu

ACKNOWLEDGMENTS

I will always be grateful to my advisor Dr. Mark Muesing, for his mentorship and support throughout my graduate education. He is an excellent intellectual and technical advisor, and a good friend. His genuine interest in the pursuit of scientific knowledge will always be inspirational.

I thank the members of my thesis committee: Drs. Brian Chait, Mike Rout, and David Ho, who have given me tremendous support. Without their advice, criticism and encouragement, the work presented here would not be possible. It has been my privilege to have them guide me. I also want to thank my external examiner Dr. Steven Goff, for kindly taking time from his busy schedule to lend his expertise to this endeavor.

I am grateful for the wonderful friendship and working relationship I experienced with Ileana Cristea, who introduced me to the world of mass spectrometry.

I thank the past and current members of the Muesing lab for making my time fulfilling. I have learned so much from these amazing and unique people.

I deeply appreciate the friendly staff in the PhD program at the Rockefeller University and my colleagues at Aaron Diamond AIDS Research Center for providing an excellent environment for me to live and study.

I thank my family and friends, who I can always count on, for their unconditional love and support over the years. I am grateful to people who have made my time in New York a wonderful experience. My special thanks goes to Eyal, for his music, companionship and love. His constant encouragement and support helped me go through the tough time, especially towards the completion of my thesis.

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

ABBREVIATIONS

AIDS – Acquired Immunodeficiency Syndrome

Amp^r – Ampicillin Resistance

AZT – Zidovudine

Bsd^r – Blastocidin D resistance gene

CA - Capsid

CAT – Chloramphenicol Acetyltransferase Resistance gene

cDNA – complementary DNA

CKII – Casein Kinase II

DMEM - Dulbecco's Modified Eagle's Medium

DNA-PK – DNA Dependent Protein Kinase

DTT – Dithiothreitol

EGFP – Enhanced Green Fluorescent Protein

FACS – Fluorescence Activated Cell Sorting

FBS – Fetal Bovine Serum

FRET – Fluorescent Resonance Energy Transfer

GFP – Green Fluorescent Protein

GHOST – LTR-GFP Reporter Containing Human Osteosarcoma Cell Line

GST – Glutathione S-transferase

HA – Hemagglutinin

HCMV – Human Cytomegalovirus

HCV – Hepatitis C Virus

HEK – Human embryonic kidney

HIV-1 – Human Immunodeficiency Virus Type 1

His – Histidine

i-DIRT – Isotopic Differentiation of Interactions as Random or Targeted

IN – Integrase

INI1 – Integrase Interactor 1

IPTG - Isopropyl-β-D-1-thiogalactopyranoside

Kan^r – Kanamycin Resistance Gene

LEDGF/p75 – Lens Epithelium-derived Growth Factor p75

LTR – Long Terminal Repeat

MA - Matrix

MAb – Monoclonal Antibody

MALDI – Matrix-Assisted Laser Desorption Ionization

MBP – Maltose Binding Protein

MCS – Multiple Cloning Site

m.o.i – Multiplicity of Infection

MS – Mass Spectrometry

MS/MS – Tandem Mass Spectrometry

NPC – Nuclear Pore Complex

PBS – Phosphate Buffered Saline

PCR – Polymerase Chain Reaction

PEI – Polyethylenimine

Pen/Strep – Penicillin/Streptomycin

PTM – Post Translational Modification

PVDF – Polyvinylidene Difluoride

RHA – RNA Helicase A

RLU – Relative Luminescence Unit

RRE – Rev Response Element

RT – Reverse Transcription or Transcriptase

SCF – Skp1-cullin-F-Box

ScFv – single-chain Fragment V

SINV – Sindbis Virus

SDS – Sodium Dodecyl Sulfate

SH3 – Src Homology-3

Sso – *Sulfolobus solfataricus*

TAR – Transactivation Response Element

vCCC – covalently closed circular viral DNA

Vif – Viral Infectivity Factor

VLP – Virus Like Particle

VPOI – Viral Protein of Interest

Vpr – Viral Protein R

VSV-G – Vesicular Stomatitis Virus G Protein

Chapter 1: Introduction to the Interaction of HIV-1 with Host Factors

Human immunodeficiency virus type 1 (HIV-1) is the causative agent of acquired immunodeficiency syndrome (AIDS) in human. Given the limited genetic capacity (10^4 bp) of HIV-1 and the complex molecular events occurring during infection, it is likely that the virus interacts with an extensive variety of cellular factors to complete its passage through the cell. To date, a handful of virus-assisting host proteins have been identified and well characterized for their mechanism of action, but many questions remain unanswered (Goff, 2007). Here we endeavor to identify the factors that interact directly with the HIV-1 machinery during viral replication using a system in which viruses have been molecularly engineered to incorporate an extremely potent immunologic peptide tag into the targeted HIV-1 proteins for investigation. As these engineered viruses were generated through a selective process based on replication competence in culture, each of the tagged viral proteins is likely to undergo the same interactions encountered by the wild type virus. We believe that this system affords us a more authentic view of both the transient and stable molecular interactions encountered during the natural course of HIV infection and therefore can be used for recovery of the tagged viral proteins from the infected culture. Indeed, using this set of independently tagged replication-competent derivatives we have recovered host proteins previously unknown to be in complex with any viral protein. Success of this work rests upon a novel methodology that preserves protein-protein interactions during cellular

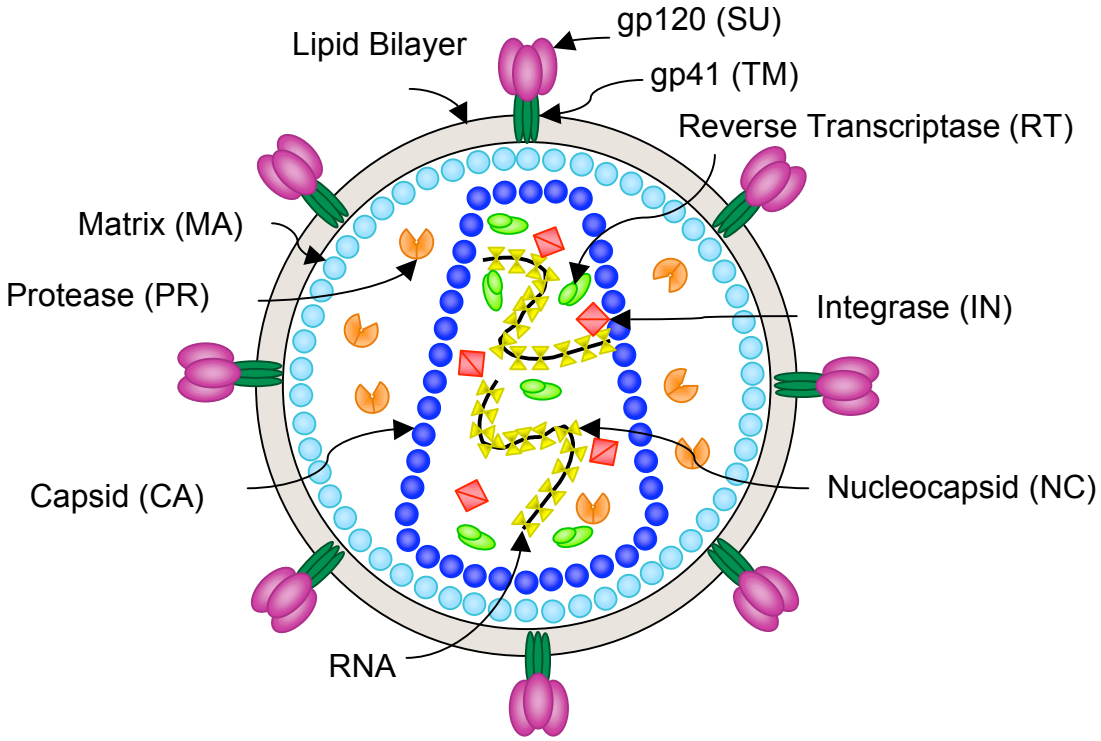
disruption and lysis that allows immunoaffinity purification of the tagged viral protein and access to its associated host factors. Elucidation of the interplay between the virus and host during HIV-1 infection would be a great asset not only for detailed insights into basic strategies employed by the virus for its propagation and survival but also to provide the rationale for the intervention of host pathways usurped by the virus as it passes through the infected cell.

1.1 HIV-1 Genome and Replication Cycle

HIV-1 belongs to a large and diverse family of enveloped diploid RNA viruses, the retroviridae, that can reverse transcribe their virion RNA into linear double stranded DNA and subsequently integrate into the host genome. Retroviruses are further subdivided into seven groups defined by evolutionary relatedness (Coffin, 1992). HIV-1 is a representative example of the lentiviruses genus, the name of which is derived from the Latin (*lenti*-, slow) characterized by a long incubation period and their chronic, persistent infections that can cause disease by killing or inducing function loss of specific cells and tissues. HIV-1 virions are about 110 nm in diameter (Darlix et al., 1995). The mature viral particle is composed of an inner core surrounded by an outer envelope formed of a lipid bilayer with cellular origin incorporating and displaying viral glycoproteins. The inner core is formed by an outer shell of capsid protein molecules (CA protein) surrounding the dimeric RNA genome in close association with about 2000 molecules of nucleocapsid protein (NC protein), each virion containing about 100 molecules of reverse transcriptase (RT) and integrase (IN) (Figure 1.1A).

HIV-1 has two copies of a single stranded, positive-sense 9.2 kilobase (kb) RNA genome. It encodes nine open reading frames, capable of producing fifteen proteins after proteolytic cleavage of the Gag, Gag-Pol and Env polyproteins (Figure 1.1B). The *gag, pol* and *env* open reading frames are common to all retroviruses. Gag (55 kDa) is processed into the mature matrix (MA, p17), capsid (CA, p24), nucleocapsid (NC, p9) and p6. They are the structural components of the internal virion core. The three Pol proteins, protease (PR, p10) for proteolytic cleavage, reverse transcriptase (RT, p66/51) containing both DNA polymerase and RNase H activity and integrase (IN) mediating integration into the host genome, all have enzymatic functions. The surface glycoprotein (SU, gp120) and the transmembrane subunit (TM, gp41) encoded by *env* form a complex that interacts with its specific cellular receptors and ultimately leads to fusion and entry of the virus into the cell. In addition to these conserved genes products, HIV-1 encodes six unique accessory proteins. Among them, Vif, Vpr and Nef are virion associated and modulate infectivity, Tat and Rev regulate viral gene expression and Vpu participates in viral assembly.

A.



B.

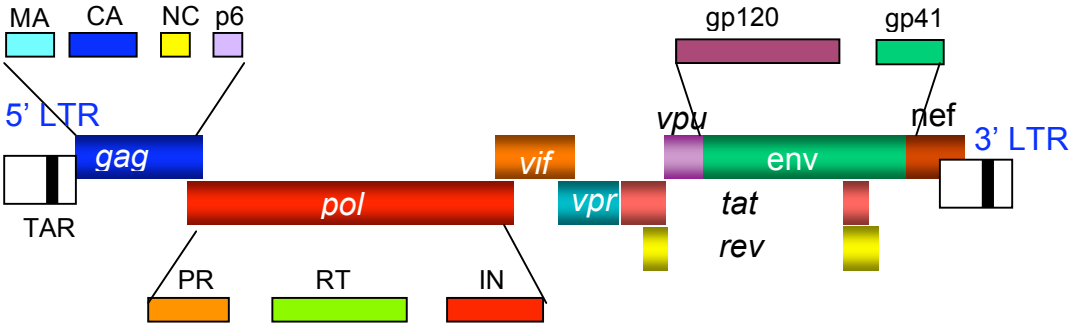


Figure 1.1 HIV-1 structure and genetic organization. Schematic representations of **(A)** cross section of HIV-1 particle, [modified from (Coffin et al., 1997)], and **(B)** the HIV-1 genome. Different reading frames are indicated by vertical displacement of the coding region.

The HIV-1 life cycle follows the general pattern of retroviral infection, but is endowed with some unique features. It can be considered as a sequence of discrete steps as outlined in Figure 1.2 (Freed, 1998). The viral surface envelope glycoprotein binds to its receptor CD4 and provides a contact surface for direct interactions with one of its associated co-receptors, either CCR5 or CXCR4. The receptor/co-receptor engagement triggers viral-host membrane fusion between the lipid bilayer of the virion and the host cell plasma membrane, releasing the viral core into the cytoplasm. Although this step is very much uncharacterized, viral uncoating is thought to occur next releasing an organized subviral composite, operationally defined as the reverse transcriptase complex (RTC) or preintegration complex (PIC), into the cytoplasm. The RTC/PIC is a high-molecular-weight assembly of several viral, and possibly cellular proteins, in combination with the two strands of nucleic acids that comprise the diploid viral RNA genome. It is in the context of the RTC/PIC that the viral RNAs are copied and converted by the activity of reverse transcriptase (RT) into a full-length linear DNA copy, the genetic information now a mosaic of that originally encoded by each of the viral genomic RNAs. Following an additional catalytic step (3' processing) in which 2 nucleotides at both ends of the viral cDNA are "trimmed" by the action of the viral integrase protein (IN) yielding recessed 3' ends, the PIC (or a subset of its original components) is transported across the nuclear membrane. In the nucleus, the viral DNA, already "pre-activated" by the trimming reaction for insertion into the host chromosome, is then recombined into cellular

DNA by another catalytic step, the strand-transfer reaction, also performed by the IN protein. The gaps left after the ligation of 3'-ends of the viral DNA to the 5'-ends of the staggered cleavage in cellular DNA are then repaired, the recombinant joint restored to its original state by the DNA repair activities of the host. Host cell mediated circularization of viral DNA leads to the formation of 1-LTR and 2-LTR circles, this event also coordinated by host enzymatic functions. The integrated viral DNA, also known as provirus, is now competent for transcription, generating the various sub-genomic RNA transcripts encoding the viral regulatory and structural gene products as well as the full-length viral RNA that is either translated to produce the large Gag-Pol polyprotein or directly packaged into the budding virion. The cellular plasma membrane is the site of virion formation and release. Here, the heavily glycosylated viral envelope protein and other structural gene products are targeted for virion incorporation along with the two new copies of viral genomic RNA. Finally, and soon after encapsidation and budding, an increase in the local concentration of the viral protease (PR) promotes its dimerization and activates PR for proteolysis. The reactive PR now cleaves the Gag-Pol polyprotein into its constituent processed products to generate the mature virion reinstating the infectious cycle.

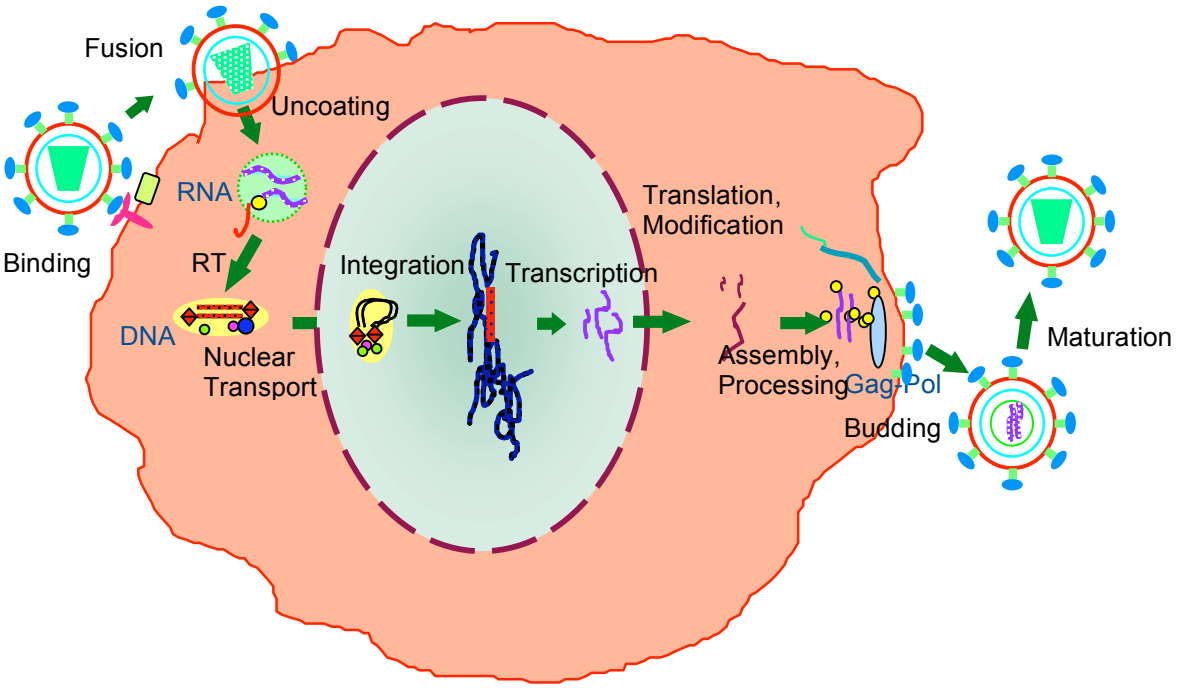


Figure 1.2 General features of the HIV-1 replication cycle. The early phase (left portion of the diagram) begins with CD4 recognition and involves events up to and including integration of the proviral DNA, and the late phase includes all events from transcription of the integrated DNA to virus budding and maturation.

As shown above, HIV-1 has a relatively small complement of viral proteins. Thus, to pass through each step of the viral life cycle, it is likely that the virus depends heavily upon the host cell for additional factors for the completion of the various molecular events required for sustained infection. In addition to the receptors and cellular transcription factors that are required for retroviral replication, it has been hypothesized that viral proteins must interact with many other cell components, making extensive use of the available host cellular machinery (Coffin et al., 1997). Moreover, it is a reasonable assumption that each of the viral proteins must perform multiple functions at several discrete points within the viral life cycle, each precisely tuned to encompass their multifunctional roles in the very limited space of the HIV-1 proteome. Discoveries like this have provided an initial glimpse into the molecular complexity at the viral-host interface. However, while most of this work has relied on the detection of direct interactions between one viral protein and its immediate cellular binding partner, the isolation and identification of the set of host factors comprising multi-component protein complexes and recruited for specific viral functions has remained elusive.

1.2 Methods for the Study of Virus-Host Interactions

Cellular co-factors for viral replication and pathogenesis have been subjects of intense investigation. Several genetic and biochemical methodologies have been employed to identify cellular genes or gene products essential to viral infection. Chemical mutagenesis has been used to induce frameshift mutations in cellular

coding regions. Mutant cell lines resistant to viral replication can be easily recovered and subsequently analyzed for interrupted gene functions (Bruce et al., 2005; Gao and Goff, 1999). Chemical mutagenesis is a simple and straightforward approach to select cell lines resistant to viral infection. However, because of the nature of chemical mutagenesis, the altered gene(s) cannot be readily identified. This fact, coupled with the possibility of altering several genes in parallel, makes it extremely difficult to identify the mutant gene(s) responsible for the associated phenotype. Another approach, insertional mutagenesis, uses retroviral integration as an insertional mutagen to randomly inactivate host genes (Joyner et al., 1989; Zheng and Hughes, 1999). This method has the advantage that the identity of the gene can be determined and the creation of a large library of independent insertional events can be readily obtained. However, this approach inactivates only one homologue and therefore can identify only those genes whose corresponding gene products are amenable to *trans*-dominant interference or can be detected by their functional haplo-insufficiency. A third method that has had success is transduction of a gene required for infection from a cell line sensitive to infection to cells that are refractory to infection. As the opposite of the insertional mutagenesis, this approach adds back positive factors of infection. For example, this technique has been very successful to recover cellular genes encoding restriction factors, host receptors or co-receptors required for viral entry (Evans et al., 2007). Transduction of a resistant genotype can also be used, especially for the study of restriction factors for viral infection (Sheehy et al., 2002; Stremlau et al., 2004). Typically, a cDNA library is prepared

from cells that are resistant to viral infection. Closely related sensitive cells are then transduced with a retroviral vector library and used to recover cells now refractory to productive infection. The cells resistant to infection can be recovered by their selected survival after an infection challenge with a virus that results in cell death, a subset of surviving cells harboring the transduced gene conferring resistance. Fluorescent screening methods have also been used (Stremlau et al., 2004). However, in addition to the difficulty of making a representative complementary DNA (cDNA) library for factors limiting in the population, these approaches will succeed only when one gene product is required for its positive or negative effects on viral infection.

Antisense library approaches and rescue methods have been developed to quickly identify antisense cDNAs responsible for a given phenotype (Deiss and Kimchi, 1991; Hannon et al., 1999). Interfering RNA (RNAi) libraries have also been used. For example, double stranded RNA, termed small interfering RNA (siRNA), homologous to a cellular gene can mediate sequence-specific degradation of the cognate mRNA upon its introduction into cells, leading to epigenetic gene inactivation (Elbashir et al., 2001). In recent years, it is clear that the use of siRNA libraries is much more efficient than antisense RNAs in knocking down gene expression and not surprisingly, this methodology has attracted a great deal of attention as a powerful tool for functional identification of genes (Vanhecke and Janitz, 2005). However, there are still some drawbacks. Studies has shown that sequence identities of as few as 11 to 12 nucleotides

between an interfering RNA and an unrelated messenger RNA (mRNA) may be sufficient for interference to occur (Jackson et al., 2003). If this is true, cross-reactivity, referred to as the 'off-target effect', could be a substantial problem (Ma et al., 2006). Although RNA interference targeting is more efficient than the antisense RNA approach, it is clear—as in the case of worms or flies—that the expression of different genes in mammalian cells are diminished with a wide range of efficiencies. The potential for a non-specific interferon response to siRNA expression in mammalian cells might also be problematic using some genetic screens (Scacheri et al., 2004).

A commonly used biochemical approach to identify co-factors for viral infection and/or replication is based on the macromolecular interactions between an ectopically expressed viral protein and its reactive host cell protein(s). In these studies, a vector is engineered to express an affinity tag fused to the N- or C-terminus of a viral protein of interest (VPOI). Upon high-level expression of the VPOI, any proteins associated with it intracellularly can be affinity purified using the fused tag and identified by techniques such as mass spectrometry or immunoblotting. Many cellular factors have been identified for various VPOI by this means (Cherepanov et al., 2003; Mulder et al., 2002; Schrofelbauer et al., 2007). However, since the targeted protein is highly expressed, this conventional and straightforward method might not necessarily represent the physiological conditions occurring during infection and may not suitable for studying the dynamics of the virus-host protein interaction.

In addition to the above methods, host proteins that directly interact with viral gene products have been identified using the yeast two-hybrid system (Kalpana et al., 1994; VerPlank et al., 2001). However, this approach requires tight interactions that take place in the yeast fungi nucleus, and is notorious for producing false positives. Moreover, this approach is not optimized for the study of most viral proteins in their natural mammalian intracellular environments.

1.3 The Interaction of HIV-1 with Host Factors

Much effort has been focused on understanding the interactions between HIV-1 and its host cells. Identification and characterization of the different cellular partners utilized by each viral protein would be invaluable in efforts to gain an intimate understanding the biology of the virus and by extension, and to expand the potential for the discovery of new antiviral targets. To date, various biochemical and molecular techniques have been used to reveal the identity of many cellular factors required for sustained propagation of HIV-1 in target cells. Furthermore, research on HIV-1 as well as many other retroviruses has shown that both viral and cellular factors regulate the progression of the viral life cycle. A list of cellular co-factors for HIV-1 characterized over the past few years and the methods used to identify them are summarized in Table 1.

Table 1. Examples of host factors that interact with HIV-1.

Viral Protein	Host Protein	Methods of Identification	Function
MA	AP3	Yeast two hybrid	Gag trafficking, virion assembly
	HO3	Yeast two hybrid	Increasing infectivity
CA	Cyclophilin A	Yeast two hybrid, *in vitro* binding	Uncoating, RT initiation
	TRIM5α	Gene transduction	Restriction factor, degradation of CA? Uncoating?
NC (p12)	APOBEC3G	Co-IP	Hypermutation, genome degradation
p6	Tsg101	Yeast two-hybrid	Budding (ESCORT recruiting)
	AIP1/ALIX	Isolation in virus like particle (VLP)	Budding (ESCORTIII recruiting)
Gag	HP68	Co-IP	Assembly
	LysRS (KARS)	Co-IP, *in vitro* reaction	tRNA packaging
	AP-2	Affinity purification	Viral budding and release
	Staufen	Co-IP	RNA packaging
	RHA	Co-IP	Particle production, RT
IN	INI1	Yeast two-hybrid	Viral assembly, nuclear import
	LEDGF	Co-IP	Viral DNA integration, chromatin tethering
	Rad18	Co-IP	DNA repair in integration
	Ku80, DNA-PK, Ku70	PIC purification	DNA repair, anti-apoptosis
	EED	Yeast two-hybrid	Import, PIC targeting
	HSP60	Affinity purification	Catalytic
	Importin 7	*In vivo* co-IP	Nuclear import
	p300	Co-IP	IN acetylation
	BAF	Co-IP	Chromatin targeting
SU (gp120)	CD4	Neutralization	Receptor for viral binding
TM (gp41)	CXCR4, CCR5	Gene transduction	Co-receptors, membrane fusion
Vif	APOBEC3G	Gene trans-duction, Co-IP	Hypermutation, genome degradation
	Cul5, Elongins B and C, Rbx1	IP/MS during infection	APOBEC3G degradation
Vpr	RIP/VpvBP	Co-IP	Cytoplasmic retention of Vpr
	DDB1	Co-IP	G2 arrest

Rev	Crm1, Rab/hRIP	Yeast two hybrid	vRNA nuclear export
	p32	Yeast two hybrid	Post-transcriptional block
Tat	Cyclin T1, cdk9	Co-IP	Transcriptional transactivation
Nef	ASK1	*In vitro* affinity binding	Signaling intermediate in apoptosis pathway
	P13K	Co-IP	blocks HIV induced apoptosis
	PACS-1	Antisense	MHC I down-regulating

Integrase (IN) is an essential viral protein that performs multiple tasks including its cooperation in reverse transcription (Wu et al., 1999), nuclear import of the PIC (Maertens et al., 2004), HIV-1 particle production (Bouyac-Bertoia et al., 2001; Gallay et al., 1997), and as indicated by its name, integration of the viral DNA into the host genome (LaFemina et al., 1992; Wiskerchen and Muesing, 1995a). In order to mediate these diverse processes, integrase is believed to associate with a number of cellular factors to exploit their functions. Among the binding partners, the integrase interactor 1 INI1/hSNF5 is a component of the chromatin remodeling SNF-SWI complex and has been found to activate IN DNA-joining activity *in vitro* (Kalpana et al., 1994). However, there is no strong evidence for involvement of INI1 during the *in vivo* integration process. INI1's binding seems to be required for late events in the viral life cycle via an unknown mechanism(s) (Yung et al., 2001). Given the timing of its activity, INI1 may interact with IN within the context of the Gag-Pol polyprotein. Alternatively, given the observed cellular relocalization of INI1 during HIV-1 infection, INI1 might play a role for virus-associated INI1 and/or cellular INI1 during the early steps of HIV-1 infection, such as PIC nuclear import (Turelli et al., 2001). Although INI1 has been hypothesized to play a part in integration-site selection or HIV-1 transcription, no experimental evidence has been produced to support this notion. On the other hand, the human lens epithelium-derived growth factor/transcription co-activator p75 (LEDGF/p75) has been identified as a new IN binding partner by co-immunoprecipitation (Cherepanov et al., 2003) and confirmed by yeast two-hybrid system (Emiliani et al., 2005). Current studies

suggest a role of LEDGF for various processes including the nuclear entry of the PIC, tethering IN to chromatin or a function in the integration reaction itself (Emiliani et al., 2005; Llano et al., 2006; Llano et al., 2004). Both INI1 and LEDGF specifically interact with lentiviral IN, not with integrases from the distantly related oncoretroviral family. This may suggest that their functional significance is restricted to lentiviral replication (Llano et al., 2004; Yung et al., 2004). Components of the DNA damage response system including DNA-PK, ATM and Ku80, have also been reported to interact with IN (Li et al., 2001). These proteins have been shown to be indirectly involved in the integration reaction, but might be required to induce the post-integration DNA repair systems that are responsible for filling in the gaps and sealing the nicks that are left after integration at the sites of viral integration (Ariumi et al., 2005; Goff, 2007; Smith and Daniel, 2006).

A notable example of exploring a cellular factor that involved in viral life cycle is the elegant study on the viral accessory gene *vif*. The vif open reading frame is present in all lentiviruses except equine infectious anemia virus and is required for viral replication and pathogenicity *in vivo* (Simon et al., 1998). In the absence of Vif, HIV-1 virions that are produced from non-permissive primary T lymphocytes and certain T cell lines, are defective, and cannot initiate productive infection. HIV-1 Vif is essential for viral evasion of the host antiviral factor APOBEC3G, which is present in Vif-nonpermissive cells and has been identified as a potent mediator of anti-HIV-1 activity (Sheehy et al., 2002). APOBEC3G

belongs to a family of proteins that have cytidine deaminase activity and when packaged into HIV-1 virions, induce modification of newly synthesized minus-strand viral DNA from cytosines to uracils. The hypermutation damages the genomic integrity and results in defective particles (Zhang et al., 2003). Vif protects the viral genome by binding to APOBEC3G and inducing its ubiquitination and degradation by a Vif-Cul5-Skp1-cullin-F-Box (SCF)-like complex, thereby blocking incorporation of APOBEC3G into the viral particles (Yu et al., 2003).

Another example of the involvement of an essential host protein pertains to the export of RNAs of HIV-1. The HIV-1 Rev protein binds to the structured Rev response element (RRE) on unspliced and singly spliced viral mRNAs and targets the mRNA to the cytoplasm using the host RNA export pathway for Rev. Rev has been shown to interact with Crm1, a host cellular protein and a main member of the nucleo-cytoplasmic transport factors absolutely required to export the viral structural mRNA transcripts from the nucleus into the cytoplasm (Neville et al., 1997).

As outlined above, a number of host proteins have been found to be involved in different aspects of HIV-1 replication including viral entry, host restriction, intracellular trafficking, mRNA export and assembly. Cellular functions that are anticipated but not yet defined will include factors ranging from chromatin remodeling to chemokine receptor-mediated signal transduction. It is certain that

the scope of the host machinery usurped by the virus is unknown and several new approaches will be necessary to find those interactions acting at the molecular interface between the virus and its host cell.

1.4 Proteomic Approach for the Study of Host Proteins that Interact with HIV-1

In our proteomic approach, we endeavor to identify factors that interact directly with the HIV-1 machinery during natural viral replication using a system in which viruses have been molecularly engineered to incorporate a potent immunological or biochemical tag. A genetic procedure has been developed for the systematic and comprehensive mutagenesis of individual HIV-1 gene segments (*in*, *vif* and *env*) in an attempt to introduce a unique restriction endonuclease site (a five amino acid in-frame open reading frame) into a viral gene without deleterious effect to the function of the associated gene product in the context of cycling viral replication. Transposon mutagenesis was used as a method to produce insertional libraries within specific HIV-1 gene segments. This protocol relies on the random introduction of a small DNA fragment encoding an efficiently recognized tag throughout a viral protein-coding segment of HIV-1. This step creates a large collection of uniquely marked bacterial DNA clones, each of which contain a small number of extra codons at different positions. The plasmid DNA from the library can be used to create a diverse viral stock and that is subsequently used to initiate HIV-1 infection of susceptible cells. Following a

stringent step for replication competency *in vivo*, viral DNA persisting after several rounds of viral replication and represented by low molecular weight extrachromosomal circular DNAs can be reintroduced back into *E. coli*, where they can be clonally purified and amplified. In turn, homogeneous viral stocks prepared from the plasmid DNA clones can be used to generate a panel of viruses each uniquely tagged with respect to the location of the insertion. Because of the rigorous step of biological selection, viruses that survive through multiple passages can accommodate a tag within a specific location within the HIV proteome without loss of biological "fitness" as revealed by their robust replication.

The ability to introduce a small peptide into a viral protein without obvious effect can be used as a mapping tool to locate a site(s) within the secondary, tertiary and quaternary structure of the protein that can accommodate foreign amino acids. Segments of protein sequence located at the surface of the three-dimensional structure or in connector regions frequently tolerate insertions of a few amino acids and remain functional, while segments that are buried, or part of an active site, are generally intolerant of such insertions (Goff and Prasad, 1991). In principle (and in practice), selected sites within proteins are usually amenable to the incorporation of even larger numbers of amino acids yet have the ability to remain functional. Here, DNA that codes for an efficient epitope or affinity tag is subsequently introduced into the HIV-1 genome within the viral gene corresponding to the preferred site of insertion.

Following confirmation of replication competency, the epitope-tagged virus can be utilized for the quantitative recovery of the viral protein. The modified viral proteins expressed in infected cells are expected to form the same physiological complexes as their natural counterparts and because of the high affinity epitope tag, each modified protein can be purified, often associated with its binding partner(s) or complex(es) that it relies on to performs its functions during viral replication through the cell. Importantly, since binding is strong, potentially transient complexes can be purified over a short period of time. A particular viral protein tagged independently at different positions will allow corroboration of any cellular binding partners identified using this procedure. The recovered proteins that are associated with each tagged viral protein can be identified using standard mass-spectrometry methods. This scheme should allow identification of the binding partner(s) for each viral protein at various times throughout the viral life cycle.

Large-scale studies of protein complexes have been carried out in yeast, in which individual proteins are tagged and used to recover associated proteins (Gavin *et al.*, Ho *et al.*, 2002). In those studies, hundreds of different proteins were tagged with a fragment of protein A or an epitope FLAG tag to create 'bait' proteins. DNAs encoding the bait proteins were then stably introduced into yeast cells. Importantly, the modified proteins form physiologically relevant complexes with other proteins that are in association with the bait protein in the context of

the living cell and thus provide a strategy to recover and identify true binding partners. Adapting this approach to HIV-1 will provide a dynamic picture of virus-host interactions not currently appreciated and contribute to the basic biological strategy employed by HIV-1 to gain control of the host cell.

Thus, the method described herein will provide a powerful tool to identify cellular proteins that associate with a given viral protein during the viral life cycle, a subset of which may be absolutely required by the virus for its livelihood but in part dispensable or compensated by the host cell. Research in this area has the potential to provide many new targets for small molecule intervention, particularly if the host factor is required for viral growth but dispensable for normal cell function. For example, individuals with a 32 base-pair deletion in the CCR5 cellular co-receptor gene are refractory to HIV-1 transmission, yet are not seemingly compromised for overall health. Consequently, small molecule inhibitors have been screened and identified against this cellular target (Strizki et al., 2001). Cellular proteins are emerging as potential targets for new antiretroviral drugs since they do not mutate as frequently as the virus and viral replication might be blocked because there are no alternative pathways. Coupled with the development of RNAi therapeutics, antiviral drugs that target cellular proteins may play major roles in combating HIV resistance.

Chapter 2: Materials and Methods

2.1 Generation of Replication-Competent HIV-1 that Incorporates a Potent Antigenic Tag

2.1.1 Random Insertion of a Unique Restriction Site (PmeI) into Proviral DNA by a Transposition Methodology

For cloning purposes, the HXb2-derived HIV-1 proviral clone, R7/3, was engineered to contain a pair of unique restriction endonuclease sites immediately flanking each of the targeted gene segments. This allows for the precise insertion of the targeted gene without disruption of the remainder of the proviral genome or plasmid backbone. The proviral clone was modified to create or eliminate certain restriction sites using overlap-PCR site-directed mutagenesis. The locations of the new restriction sites with respect to the corresponding amino acid (aa) within the viral proteins studied here are as follows in Table 2.

Table 2. The location of restriction sites with respect to their corresponding amino acid within the viral protein.

Viral Protein Segment	***5' Site***	***N-terminal amino acid****	***3' Site***	***C-terminal amino acid****
Complete Integrase (IN)	*XbaI*	*4*	*MluI*	*288 / 288*
Integrase C-Terminus (IN-CTD)	*PstI*	*215*	*MluI*	*288 / 288*
Vif	*SacII*	*51*	*BstEII*	*185 / 192*
Env (V1-V3)	*NdeI*	*62*	*NheI*	*345 / 856*

*The location of the first and last codons within the viral protein-coding segment that can be mutagenized without interruption of the adjacent restriction enzyme recognition sequence. The total number of amino acids of the complete protein is indicated in the right column after the slash. None of the mutagenic alterations that introduce a new restriction site alter the identity of the corresponding amino acid in the target viral protein.

To generate a library of transposon-mediated insertions in the individual viral gene segments, the regions of interest were subcloned into the ampicillin resistant pBluescript II (pBS) KS(-) plasmid (Stratagene) to generate pBS-KS-*in*, pBS-KS-*in(ctd),* pBS-KS-*env(v1-v3)* and pBS-KS-*vif*. These pBS-KS(-) subclones served as substrates for the *in vitro* transposition reaction using the Tn7-derived, GPS-LS linker scanning system according to the manufacturer's instructions (New England Biolabs). The transposition reactions were used to create plasmid libraries – pBS-KS-*in/*Tn, pBS-KS-*in(ctd)/*Tn, pBS-KS-*env(v1-v3)*/Tn and pBS-KS-*vif*/Tn, each containing a large number of random 1.7 kb Kan^r transposition insertion within the targeted viral gene. Greater than 10^6 recombinants were selected based on ampicillin/kanamycin double resistant phenotype they confer upon the host bacterial cell. To recover only those transpositional events that occur within the targeted gene segment, plasmid DNA from reach plasmid library was combined and the pooled DNA preparation (QIAgen) digested with the corresponding restriction enzymes, XbaI/MluI for pBS-KS-*in/*Tn, PstI/MluI for pBS-KS-*in(ctd)/*Tn, NdeI/NheI for pBS-KS-*env(v1-v3)*/Tn and SacII/BstEII for pBS-KS-*vif*/Tn. Viral gene segments that received a transposon were purified from those gene segments that did not contain a transposon based on their inherent mobility difference on agarose gels (i.e., the transposon-containing-fragments were 1.7 kb larger in size). The transposon-targeted fragments were isolated [QIAquick Gel Extraction Kit (QIAgen)] and then ligated into the appropriately restriction-site-modified R7/3 proviral vector clone. To recover proviral plasmids containing a transposon within the targeted gene segment, bacterial transformants were

selected based on their combined resistant phenotype (Amp^r-Kan^r) and the plasmid DNA from each of double resistant R7/3 library pooled and digested with PmeI. The PmeI restriction site does not appear in the parental plasmid R7/3, but flanks the ends of the inserted transposon in the Amp^r-Kan^r transformants in such a way that digestion with PmeI removes the element from the plasmid. Therefore, the PmeI-restricted, linear DNA from each of the R7/3/Tn libraries was purified, ligated to itself, retransformed back into bacteria and the number of colonies from each library determined. Each library now contains a 15-nucleotide insertion randomly and uniquely placed throughout the target viral coding region (*in, in-ctd*, *env* or *vif*). To confirm this supposition and to map the distribution and randomness of our mutagenesis, each plasmid library was digested with PmeI and a restriction enzyme lying outside and stationary with respect to the endpoints of the targeted viral gene. Plasmid DNA was again a prepared, individual library pooled and used to prepare the heterogeneously tagged viral stocks.

2.1.2 Viral Production and Infection

Human embryonic kidney (HEK) 293T cell lines were passaged at 37°C in the presence of 5% CO2 in culture medium containing Dulbecco's modified Eagle's medium (DMEM) supplemented with penicillin and streptomycin (Pen/Strep) plus 10% fetal bovine serum (FBS). To produce virus, 3 µg pooled DNA from the R7/3-PmeI libraries (*in*, *in-ctd* or *env)* was used to transfect subconfluent monolayers (30%) of HEK 293T cells grown in T-25 culture flasks using 25 µg

polyethylenimine (PEI) (Polysciences Inc.). For the R7/3-PmeI *vif* library, viral stock was made by transfecting 293-APOBEC3G-HA. This cell line is derived from HEK293 cells and stably transfected with a pcDNA expression vector for human APOBEC3G tagged with HA (Yu et al., 2003).

Following DNA transfection, the flasks were incubated at 37^{o}C for 18 hours, the transfection medium removed and replaced with 10 mL of fresh medium added back to each transfection, and the supernatant recovered and aliquoted. Viral supernatants were collected 20-24 hours later (38-42 hours post-transfection) by centrifugation at 1,000 rpm for 5 min. Viral production was determined by p24 ELISA performed using the HIV-1 p24 Antigen EIA kit (Beckman Coulter). Viruses with 10 ng of p24 produced from *in* or *env* libraries were used to initiate infection in 1 x 10^6 CEM cells, and MT2 cells were used for infection by viruses produced from *vif* proviral library. All infected cultures were propagated in T-25 flasks at 37 °C in RPMI-1640 supplemented with 10% FBS. Progress of the advancing infection was monitored visually by syncytia formation, the cell density maintained at approximately 1 x 10^6 cells per mL with 1:3 passage when necessary. For sample collection, 1 mL of the infected culture was pelleted down at 1,000 rpm for 5 minutes. Cell pellets were collected once a day over the course of 7 days post-infection, washed in phosphate buffer saline (PBS) and stored at -80^{o}C until further processing.

2.1.3 Purification of Circular Viral DNAs from CEM Infections—

Recovery of Replication-Competent, PmeI-Tagged HIV-1 Clones

The Hirt extraction procedure (Hirt, 1967) was used to recovery low molecular weight viral DNAs, among these DNAs, covalently closed 1- or 2-LTR viral DNA circles produced in infected cells were collected each day from each infected culture. Briefly, the cell pellets from infections were washed once with PBS and then resuspended in 500 µL of 10 mM Tris-HCl/10 mM EDTA. After adding 50 µL of 10% SDS and incubation for 10 min at room temperature, 140 µL of 5 M NaCl was added, the mixture was gently mixed and the insoluble fractions allowed to precipitate overnight at 4^{o}C. The supernatant was collected after centrifugation at 14,000 rpm for 30 min. The DNA was isolated by ethanol precipitation using 0.25% linear acrylamide as carrier and resuspended in 20 µL of a 10 mM Tris-HCl (8.0) buffered solution.

Five microliters of each DNA sample prepared by the Hirt extraction method were used as template for PCR. The viral coding regions (*in, env, or vif*) harboring unique PmeI insertions were amplified by 25 cycles of PCR using the appropriate primer pairs for each gene and the resulting PCR-amplified DNA fragments purified using QIAgen Gel Extraction Columns (QIAgen) and eluted in 30 µL 10 mM Tris-HCl/EDTA buffer. To determine if viral recombination had occurred via reverse transcription during any part of expansion and led to the emergence of

wild type viral DNA sequence within the targeted gene sequence, 5 µL of each PCR product was cleaved with PmeI. The percent of viral DNA that had retained or lost the PmeI site during infection was determined using 0.8% agarose gel electrophoresis. For selected samples, the purified PCR-amplified DNA was digested with the two unique restriction enzymes that flank each viral reading frame, namely, XbaI-SacII for *in*, PstI-MluI for the integrase C-terminal domain, NdeI-NheI for *env [gp120 (v1-v3)]* and SacII-BstEII for *vif*. Resulting fragments were ligated into its corresponding pBS-KS(-) vector to generate libraries with the PmeI-tagged insertion in the viral gene of interest. Since only those plasmids containing a PmeI site were linearized after PmeI digestion, the linear DNA resulting from this cleavage reaction was purified away from supercoiled DNA by electrophoresis over 6% polyacrylamide and electroelution of the linear DNA (supercoiled DNA does not enter this type of gel). Following self-ligation and bacterial transformation, plasmid DNAs with PmeI insertions were prepared from each of 10-30 individual clones and the targeted viral gene segment of each completely sequenced to determine the location of the PmeI tag. Tagged viral genes were recloned back into proviral R7/3 and individual viral stocks prepared and tested for replication competency (syncytia formation). Positive, infectious clones represent mutant HIV-1s that can tolerate a 15 bp PmeI insertion yet remain viable and as such are excellent candidates for further modifications including the addition of a 3xFLAG epitope tag at the unique PmeI site within the IN, Env and Vif proteins.

2.1.4 Construction of Replication-Competent 3xFLAG-Tagged HIV

Three commercially-prepared (IDT) synthetic oligonucleotide complimentary pairs encoding all three reading frames of the 3xFLAG epitope (NH2-GGSADYKDHDGDYKDHDIDYKDDDDKASGG-COOH) were phosphorylated at their 5' ends, annealed and the appropriate oligonucleotide pair cloned into the PmeI site of their respective replication competent PmeI-proviral clone. To prevent multiple 3xFLAG oligonucleotide insertions, the resulting plasmids were cut with BlpI. BlpI is a restriction site specified by all oligonucleotide pairs but absent from the rest of the plasmid. In addition, since BlpI recognizes an abbreviated palindromic sequence (5'…GC/TNAGC…3'), only ligation events regenerating the complete 3xFLAG DNA sequence are possible. Oligonucleotide pairs encoding several other tags such as 1xFLAG (GGSADYKDDDDKASGG), an amino acid sequence capable of biotin addition we designate, SNAGG (GGSASSLRQILDSQKMEWRSNAGGSGGV) and a biotin mimicking Strep tag (GGSAWAHPQFEKASGG) were also constructed. Proviral clones with these tags were generated using the same methodology described above. Thus, while the methods described herein focus on the 3xFLAG tagged clones, in principle, this strategy can be applied to any other tag.

The resulting tagged proviral clones were transfected individually into 293T or 293-APOBEC3G (in the case of the *vif* tagged clone) cells using PEI as

described above to produce viral stocks with a 3xFLAG in the *in*, *env* or *vif* reading frames. Viral production was quantified by p24 ELISA, and viruses corresponding to 10 ng of p24 used to infect 1 x 10^5 CEM or MT2 cells (Vif-tagged viruses) in 6-well plates. Cell counts were maintained at 1 x 10^6 cells per mL by splitting 1:3 on passage. Replication-competent clones with in-frame 3xFLAG tag insertions that resulted in a progressive and robust syncytia formation after their infection of CEM cells were used for further experimentation.

2.1.5 Concentrating Viral Particles

Viral supernatants (0.7 mL) from each 3xFLAG tagged virus were collected 7 days post-infection and filtered using 0.45 µm filters (Millipore). Viruses in cell-free media were carefully layered onto the top of 0.3 mL cushions of 20% sucrose in PBS in 1.5 mL minicentrifuge tubes. Virus was pelleted at 14,000 rpm at 4^{o}C for 3 hours. Supernatant were removed by aspiration and trace amounts of liquid removed by pipetting. Viral particles remaining at the bottom of each tube were resuspended and inactivated in 20 µL of 1% Epigen in PBS at room temperature for 30 min.

2.1.6 Immunoblotting Analysis of Viral Proteins

Infected cell pellets were washed twice with PBS and suspended in RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate and 0.1% SDS) containing mixed protease inhibitors (Roche). Samples were incubated with rotation at 4^{o}C for 30 min, sonicated three times at a 10% setting

for 10 sec and then cleared by centrifugation at 14,000 rpm for 10 min at 4°C. Virion pellets and soluble portions of cellular lysates were mixed with LDS sample buffer (Invitrogen) and 10mM DTT, heated for 5 min at 95°C. Proteins were resolved by electrophoresis on 10% or 4-12% Bis-Tris polyacrylamide NuPAGE gels (Invitrogen), and the gel's contents transferred to polyvinylidene difluoride (PVDF) membranes (Millipore). Molecular weight estimates for experimental bands on immunoblots were obtained by comparison with the mobilities of *New Precision Plus Protein All Blue Standard* (Bio-Rad) or the *SeeBlue Plus 2* pre-stained standard (Invitrogen).

Membranes were subjected to immunoblot analysis by incubating with primary antibody at 4°C overnight or 1 hr at room temperature. Primary antibodies were as follows: 1 µg/mL α-FLAG mouse mAb M2 (Sigma) for 3xFLAG tagged viral protein expression; 1:1,000 6G5 integrase monoclonal antibody (Nilsen et al., 1996); 1:1,000 monoclonal α-Vif (Simon et al., 1995) or human polyclonal IgG purified from the serum of an HIV-1 positive individual (LSS) classified as a long-term nonprogressor (Cao et al., 1995). After washing the blot three times with 0.1% Tween in PBS, the membranes were incubated with secondary antibody at 1:10,000 for 1 hr at room temperature. Sheep anti-mouse horseradish peroxidase-conjugated antibody (Amersham Biosciences) was used as the secondary antibody if a murine monoclonal was used as the primary antibody. Anti-rabbit IgG was used for polyclonal primary antibodies and human sera was followed by peroxidase-conjugated anti-human antibody incubation. Proteins

were detected with the *ECL Plus Chemiluminescence System* (Amersham Biosciences) or *Immobilon Western HRP Chemiluminescent Substrate* (Millipore).

2.2 Characterization of the Replication-Competent Tagged Viruses

2.2.1 Viral Growth Curves

To assess the growth kinetics of the R7/3-IN-3xFLAG and R7/3-Env-3xFLAG viruses, these tagged viruses as well as the unmodified, control wild type, R7/3, were produced in transfected 293T cells as described earlier. $1x10^6$ CEM cells in T25 flasks were infected with viral supernatants normalized for p24 levels (10 ng). Input virus was removed after overnight incubation by washing cells twice with PBS, and infected cells maintained in fresh RPMI medium. Culture supernatants (200 µL) were collected starting at the day of infection daily for 7 days. The rate of viral growth was quantified by measuring the concentration of p24 in the culture supernatants, as determined by an HIV-1 p24 antigen EIA kit (Beckman Coulter). Infection in MT2 cells with R7/3-Vif-3xFLAG or wild type R7/3 produced from 293-APOBEC3G were similarly performed.

2.2.2 Production of VSV-G Pseudotype Viruses

Pseudotyped HIV-1 particles carrying the vesicular stomatitis virus G protein (VSV-G) were used in this study (Naldini et al., 1996; Reiser et al., 1996). The

NdeI site in the *env* gene was restricted with NdeI and Klenow polymerase used to blunt-end the site. In the complete proviral DNA, this procedure results in a frame-shift in the *env* reading frame (*env*^{minus}) and loss of functional Env protein. Pseudotyped viruses were generated by DNA co-transfection of 293T cells with 0.5 µg of a pVSV-G expression vector (pCI-VSV-G) DNA and 0.5 µg of the env^{minus} proviral plasmid derivative per well in 6-well plate format containing 2 mL of DMEM complete media. In addition, 8 µg/mL polybrene was added to each transfection. The medium was changed after 12 hr. Then, 48 hr after transfection, the medium was removed and filtered through 0.45 µm Millex-HA filter (Millipore). Virus stocks were aliquoted and stored at -80°C. Extracellular p24 levels were measured as described above.

2.2.3 One-Step Infectivity Assays

Three indicator cell lines were employed to assess viral infectivity. (1) P4R5 cells, a HeLa-CD4/LTR-*lac*Z indicator line expressing both CXCR4 and CCR5. In this cell line, the ß-galactosidase reporter gene is expressed only in infected cells (Charneau et al., 1994). (2) The GFP reporter-containing human osteosarcoma cells (HOS)-derived indicator cells, GHOST(3)/X4/R5 containing the HIV co-receptor molecules, CCR5, CXCR4 (Morner et al., 1999) and R3 and stably transfected with a Tat-dependent reporter construct consisting of the HIV-1 long terminal repeat (LTR) enhancer-promoter directing the expression of a humanized allele of the green fluorescent protein (GFP). (3) A CEMx174 derivative cell line, LuSIV, modified by stable transfection of a luciferase reporter

gene (Roos et al., 2000) activated only upon infection.

For the adherent cell lines, HeLa-P4R5 or GHOST(3)X4R5, confluent monolayers were maintained on 10-cm dishes and split 1:10 into 6-well plates the day before infection. For suspension cells, LuSIV, $1x10^5$ cells were plated into each well of a 6-well plate. 1 ng of virus was used for infection in the presence of polybrene at 8 µg/mL. Plates were incubated at 37^{o}C for 12 hours, after which cells were washed three times with PBS and the unattached viruses removed. LuSIV infections were incubated for an additional 2 days at 37^{o}C. The cells were resuspended in 2 mL of media. 100 µL of the infected culture transferred to a 96-well plate compatible with the MLX 96 Well Plate Luminometer (Dynex). 100 µL of Bright-Glo Buffer with Bright-Glo substrate (Promega) was added to each well and mixed with the cells by pipetting. After incubation at room temperature for 2 minutes, samples were processed for luminescence. The output data was analyzed using the Revelation software platform. Infectivity of various tagged viruses is expressed as percentages relative to the wild type R7/3 infectivity quotient.

A genetic assay was employed to measure the integration frequency of the viruses with tagged insertions. The *nef* region of the WT and its derivative proviral clones with PmeI, 1xFLAG or 3xFLAG insertions within the integrase gene was replaced with the dominant selectable marker, the blastocidin D gene (bsd). All integrase clones with insertions also incorporate inactivating frameshift mutations in the *vpr* and *env* genes. Vesicular stomatitis virus (VSV)-G

pseudotyped viral stocks were prepared by co-transfecting the R7/3 BSD *vpr*^{minus} *env*^{minus} proviral vectors with pCI-VSV-G into 293T cells. Viral production was measured by p24 determination. HeLa P4R5 cells were plated in 12-well format at approximately 30% confluence and infected with 2, 20 or 200 ng of p24 in duplicate in 0.5 mL culture media for 8 hrs. The medium was then changed and cells incubated for an additional 36 hrs before being trypsinized and serially diluted in complete medium supplemented with 5 µg/mL Blasticidin-S HCl (Invitrogen). Cells were cultured for an additional 6 days before staining with a fixing solution containing crystal violet. Colonies were counted to ascertain relative integration frequencies.

2.2.4 Neutralization Assay

Neutralization assays were employed on the Env-3xFLAG tagged virus to determine the effects of 3xFLAG insertion on the ability of a known V3-directed neutralizing antibody, 447-52D (Gorny et al., 1992), and the ability of the M2 α-FLAG mAb to neutralize Env-tagged infections. Neutralization was calculated as reductions in luciferase or EGFP reporter gene expression after multiple rounds of virus replication in the dual reporter suspension cell line, 5.25.EGFP.Luc.M7 (gift from Dr. N.R. Landau, New York University School of Medicine, New York). The cell line is a genetically engineered clone of CEMx174, a fusion of 721.174 cells to the human T cell line CEM, and expresses multiple entry receptors: CD4, CXCR4, GPR15/Bob and CCR5 (Brandt et al., 2002). It also possesses Tat-responsive reporter genes for both luciferase and the green fluorescent protein

(GFP). These cells were maintained in growth medium containing puromycin (0.5 μg/mL), G418 (300 μg/mL) and hygromycin (200 μg/mL) to preserve the CCR5 and reporter-gene segments.

For neutralizing assays, 10 ng and 100 ng p24 of cell free R7/3 WT and the infectious clone with 3xFLAG tag insertion in the V2 region (R7/3-Env-3xFLAG) were incubated with different dilutions of the M2 α-FLAG (0.4-100 μg/mL) or a conventional HIV-1 neutralizing antibody, 447-52D (0.1-25 μg/mL) for 1 hr at 37^{o}C in a 24-well plate. A 100 μL suspension of susceptible cells (5x10^{5} cells/mL) was then added to each well. The virus-antibody complexes were adsorbed to 5.25 M7 cells for 12 hrs at 37^{o}C. Unbound viruses and antibodies were removed by centrifugation and washing with PBS. One well received cells plus virus in the absence of antibody as a viral control for productive infection. Another set only received cells and was used as the background control. The reverse transcriptase inhibitor, AZT (10 μM) was added to one well of the infections with no antibody as an additional control. Plates were incubated for 3 days or until approximately 10% of the cells in the virus control wells were positive for GFP expression as determined by fluorescence microscopy. At that time, 100 μL of cell suspension was transferred to a 96-well white solid plate (Costar) for measurement of luminescence by use of Bright-Glo substrate solution (Promega), in accordance with the manufacturer's instructions. Neutralization titers were considered positive when relative luminescence units (RLUs) were reduced by either 50% compared with the viral control wells without

added antibody and subtracted for the background signal.

2.2.5 Flow Cytometry

To investigate the effect of neutralizing antibodies on the 3xFLAG insertion into Env, 20 ng p24 of either R7/3 WT or Env-3xFLAG were incubated with different dilutions of M2 α-FLAG (2-100 µg/mL) or neutralizing antibody 447-52D (0.5-25 µg/mL). After incubation for 1 hr at 37^{o}C, the virus-antibody mixtures were added to semi-confluent GHOST(3)X4R5 cells in 24-well plates, and spinoculated at 2100 rpm for 1 hour. After incubation for an additional 4 hrs at 37^{o}C, unbound viruses and antibodies were aspirated and the cells washed with PBS. One well received cells plus virus in the absence of antibody as a positive control for viral infection. Another set only received cells and was used as the background control. Cells were rinsed with PBS and harvested 2 days post infection by treating with trypsin-EDTA for 30 sec, and fixed in 4% para-formaldehyde before analysis by flow cytometry (GFP fluorescence). A FACS Calibare (Becton-Dickinson) was used for analysis coupled with CellQuest software (Becton-Dickinson). Control staining was performed using uninfected cells. Results are plotted as a frequency versus the log fluorescence of the signal or its geometric mean.

2.2.6 Immunofluorescent Microscopy and Image Deconvolution

Hela P4R5 or GHOST(3)X4R5 cells were plated on glass coverslips with poly L-lysine treatment (0.01% solution, Sigma) at 20% confluency in 24 well plates.

Cells were infected on the second day with 10 ng of tagged or wild type viruses and incubated for an additional 48 hours. Alternatively, cells were plated to 30% confluency and analyzed over time. For the time course study, cells were infected with 10 ng of tagged or wild type virus at 9-hour intervals until 36 hours post-infection. Cells were then fixed for 30 minutes on ice with freshly made 3% formaldehyde/PBS (Tousimis), and permeablized with 3% formaldehyde, 0.2% Triton in PBS for 15 minutes on ice. After blocking with 1% goat serum/PBS for 15 minutes, cells were incubated for one hour at room temperature with monoclonal M2 FLAG antibody (4.8 μg/μL, Sigma) at 1:2,000 dilution in PBS containing 1% goat serum. Following the primary antibody incubation, cells were washed with PBS/0.1% Tween and incubated for an hour at room temperature with the fluorescein-conjugated secondary antibody, Alexa 594 Fluoro-coupled goat α-mouse antibody (Molecular Probes), at 1:500 dilution in PBS/1% goat serum. After three washes in PBS/0.1% Tween, cell-coated coverslips were removed from the plate and left to air dry briefly. One drop of Vecta-Shield mounting media containing 4,6-diamino-2-phenylindole (DAPI) for nuclear staining (Vector Laboratories, Burlingame, CA) was added to each microscope slide. Alternatively, nuclei were stained with 0.8 μM bis-benzimide (Sigma, Hoechst No. 33258) in PBS. The coverslips were then placed on slides, cell side facing down on top of the mounting media and sealed with nail polish.

An inverted DeltaVision optical sectioning microscope (Applied Precision) was used to collect between 10 and 20 images spaced at 0.2 μM through the slide

using standard FITC, Cy5 and rhodamine filters. Images were acquired with a 60x oil immersion objective. The images were deconvolved through 15 iterations using the DeltaVision deconvolution software (Applied Precision). 3D volume reconstructions and sections were obtained by compiling the deconvolved sections with the Softworx software (Applied Precision).

2.2.7 Pilot Immunoisolation Experiments

A small-scale Immunoprecipitation experiment was performed to investigate the efficiency of isolating 3xFLAG tagged viral proteins. Viruses were produced as described above. Infection was initiated in $1x10^6$ CEM cells in T25 flasks with viruses tagged with the 3xFLAG tag in integrase or envelope gp120 (R7/3-IN-3xFLAG or R7/3-Env-3xFLAG, respectively) equivalent to 10 µg p24. Cells were harvested by centrifugation at 1000 rpm for 5 min six days post infection or when most of the cells in the culture were involved in syncytia formation. After washing twice with PBS, cell pellets were resuspended in 500 µL modified RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate and 0.1% SDS) in the presence of Complete™ protease inhibitor (Roche). After 30 min incubation at 4°C, lysates were sonicated with a Sonic Dismembrator 500 (Fisher Scientific) and cleared by 10 min of centrifugation at 14,000 rpm at 4°C. Cell lysates were incubated with 0.2-0.5 mg of Epoxy 270 Dynabeads (Invitrogen) conjugated with M2 α-FLAG (Sigma) for 1 hr at 4°C with gentle rotation. After 6 washes with modified RIPA buffer, proteins were eluted with 20 µL of 1x LDS sample buffer (Invitrogen). For western blot analysis, eluted

proteins as well as equal percentages of the total lysate and flow through were separated by electrophoresis on Novex polyacrylamide gels (Invitrogen), and then transferred to polyvinylidene difluoride membranes (Amersham Biosciences). The membranes were incubated with 1 µg/mL M2 α-FLAG mAb for 3xFLAG-tagged viral protein detection. A sheep α-mouse horseradish peroxidase-conjugated antibody (Amersham Biosciences) was utilized as secondary antibody and the reactive proteins visualized using the ECL-Plus Chemiluminescence System (Amersham).

2.3 Identification of Cellular Proteins Interacting with Tagged Viral Proteins During Infection

2.3.1 Virus Production

For the integrase or Env gp120 3xFLAG tagged viruses, viral stocks were produced by transfecting HEK293T cells in T75 flasks with 8 µg proviral DNA using 70 µg PEI. After changing media with 12 mL fresh RPMI 18 hrs later, supernatants were harvested 40 hours post transfection at 1,000 rpm for 5 min, aliquoted and frozen at –70^{o}C. Viral production was quantified by p24 ELISA as described above. For virus with 3xFLAG tag in Vif, viral stock was generated by transfecting 293-APOBEC3G-HA under the same conditions.

2.3.2 Infection and Culture Expansion

Human T-cell lines, CEM (Vif-permissive) and MT2 or HuT78 (both Vif-restricted cell types), were maintained in RPMI 1640 medium (Invitrogen) supplemented with 10% FBS and Pen/Strep. For 3xFLAG tagged IN or Env viruses, CEM cells from 40 ml of logarithmically growing culture were concentrated by spinning at 1,000 rpm for 5 min. Viral supernatants of 5 µg of p24 were added to cell pellets and additional RPMI were added to give 10 mL of total volume. The virus-cell mixture was spinoculated at 2,300 rpm at 20^{o}C for 1 hour. An additional 30 mL of fresh RPMI was added to the culture and incubated at 37^{o}C. Input virus was removed by washing the cells three times with PBS 18 hours post-infection. Infected cells were resuspended in 120 mL of fresh RPMI in T175 flasks. Infections were expanded 1:3 every other day in T175 flasks. An additional 5x10^{7} uninfected CEM cells were added to each flask when they were “blossoming” with microscopic syncytia formation. Infected cells were harvested six days post infection. In each infection, there was approximately 1 L of suspension culture collected. For the Vif 3xFLAG tagged viruses, human T-cell lymphoma HuT78 and MT2 cells were used as host cells for infection. The rest of the steps for these cell types were similar to the procedure outlined above for CEM cells.

2.3.3 Conjugation of Magnetic Beads with Antibody

A suitable amount (20-60 mg) of M-270 Epoxy Dynabeads (Invitrogen) were washed three times with 1 ml of 0.1 M sodium phosphate buffer, pH 7.4. During the second wash, the beads were incubated with shaking for 10 min at room

temperature. After the third wash, cells were resuspended with M2 α-FLAG mAb (Sigma) at the ratio of 10-20 µg of antibody per mg of beads. Sodium phosphate buffer (0.1 M) was added to bring the total volume to 13.3 µL/mg of beads. After gentle mixing, 6.7 µL/mg beads of 3 M ammonium sulfate was added to give a total volume of 20 µL/mg of beads. The mixture was incubated at 30 °C overnight with gentle rotation. Following conjugation, the solution was removed after applying a magnetic field. 0.1 M sodium phosphate buffer was immediately added to the beads. Subsequently, the coupled magnetic beads were washed in the following order: 100 mM glycine-HCl pH 2.5, 10 mM Tris pH 8.8, freshly prepared 100 mM triethylamine solution, four washes with PBS, one 15-min wash with 0.5% Triton X-100 in PBS and one final wash with PBS. Beads were stored in PBS/0.02% NaN_3 at 0.1 mg beads/µL at 4 °C up to a month.

2.3.4 Cell Disruption and Cell Lysis

Approximately 1 L of CEM cells infected with R7/3 with 3xFLAG insertion in either IN or Env, and HuT78 infected with viruses 3xFLAG tagged in Vif were harvested six days post infection by centrifugation at 1,000 rpm. In addition, viral particles from 1 mL of the cell-free supernatant were concentrated following the procedures described above. Cells were disrupted and lysed based on the successful development of a cryogenic method (Cristea et al., 2006; Cristea et al., 2005). After washing the cell pellets twice with PBS, the cell pellets were weighed and resuspended at the ratio of 300 µL/g of cells in 20mM HEPES, pH 7.5, 1.2% (w/v) polyvinyl-pyrrolidone (PVP, Sigma) with a protease inhibitor

cocktail (Roche) and integrase inhibitor Merck-0812 at 2 μM (Merck). The cell mixture was dripped into liquid nitrogen by careful pipetting and then snap frozen to form small pellets of uniform size. Lysis buffer was formed into frozen pellets as well. Lysis buffer is 20mM K-HEPES, pH 7.4, 110mM KOAc, 2mM $MgCl_2$, 0.1% Triton, 0.5% deoxycholate, 300mM NaCl, 25 units /mL DNase I (Roche) supplemented with protease inhibitor (Roche) and HIV-1 integrase inhibitor. Snap-frozen pellets were also made by pipetting the complete lysis buffer into liquid nitrogen. Pellets from each viral infection was combined with lysis pellets formed from 2 mL lysis buffer and loaded into a 25 mL metal grinding jar and topped with a 20 mm tungsten carbide grinding ball. Cells were disrupted by cryogenic grinding using Retsch MM 301 Mixer Mill (Retsch). The grinding was completed in 6 cycles at 30 Hz for 3 min. The jars were recooled in liquid nitrogen between cycles to prevent sample thawing. After grinding, the jars were kept closed on ice until the grounded cells and lysis buffer thawed and mixed in the liquid phase. After a brief spin of the contents of the jar in a preparative centrifuge, the lysates were carefully collected. Additional lysis buffer was added to give 6 mL lysis buffer in total per g of pelleted infected cells and used to rinse the jar of any trace amount of lysate. Cell lysates were slowly rotated at 4°C for 15 min, vortexed vigorously three times for 30s followed by another incubation with rotation for 10 min followed by an additional vortex. Cell lysates were then centrifuged for 10 min at 3600 rpm at 4°C. The supernatant fraction was transferred to a new tube and used for the affinity purification experiments. To test extraction efficiency, the pellet formed after centrifugation was resuspended

in 1 mL 1x LDS sample buffer and sonicated with a Sonic Dismembrator 500 (Fisher Scientific), heated at 90^{o}C for 10 min and cleared by 10 min centrifugation at 14,000 rpm.

2.3.5 Immunoaffinity Purification

Suitable amounts of α-FLAG conjugated magnetic beads (7 mg beads per gram of infected cells) were washed three times with 1 mL lysis buffer. After the last wash, the beads were divided into 200 µL aliquots in lysis buffer and added to the soluble fraction of cell lysate from each infection. Cell lysate and beads were incubated with gentle rotation at 4 °C for 1 hour. The magnetic beads were concentrated by applying a magnet field. The column flow-through was collected to test the efficiency of protein recovery. The beads were then washed six times with 1 mL lysis buffer. The isolated protein complexes were eluted from the beads by shaking gently for 20 min at room temperature with 500 µL of a fresh 0.5 N NH4OH, 0.5 mM EDTA solution. Samples were frozen on dry ice and dried in a speed vacuum overnight. The lyophilized elutants were resuspended in 30 µL 1x LDS-PAGE sample buffer with reducing agent (Invitrogen). After heating at 70^{o}C for 10 min with agitation, 100 mM iodoacetamide was added to the samples and incubated at room temperature in the dark for 30 min to alkylate the reduced cysteines.

Ten percent of each LDS sample buffer elutant was saved for Western Blot analysis. The rest of the samples were separated on a 4–12% NuPAGE Novex

Bis-Tris precast electrophoresis gel (Invitrogen). The gel was fixed with 16% methanol and 10% acetic acid in water for 10 min and stained with Gel-Code Blue Stain reagent (Pierce) according to the manufacturer's fast staining protocol. Briefly, the gel was rinsed four times and heated in a microwave twice. Staining solution was added to cover the gel and heated until boiling, followed by incubation at room temperature for 30 min with gentle shaking. Another staining was performed to enhance the signals of protein bands. The gel was destained by several rinses in double distilled water (ddH_2O) and shaking for 20 min in ddH_2O. The destained gel was stored in ddH_2O at 4^oC until processing.

2.3.6 Mass Spectrometry

To identify the proteins in the gel, each lane of elution was cut from the gel and excised into 24-26 1 mm wide slices. Two or three adjacent slices were combined and then diced into $1mm^3$ cubes. They were destained in a solution of 50 mM ammonium bicarbonate and 50% acetonitrile with agitation at 4^oC. After discarding the solution, the gel pieces were dehydrated in acetonitrile, and in-gel-digested with 12.5 ng/μL sequencing grade modified trypsin (Promega) overnight at 37^oC. The resulting peptides were extracted overnight at 4^oC using reverse phase resin POROS 20 R2 (Applied Biosystems) at the concentration of approximately 1.25 g resin/L. After several washes in 0.1% (v/v) trifluoroacetic acid (TFA), proteins were co-crystallized with 50% saturated α-cyano-4-hydroxycinnamic acid (4-HCCA) in 70% (v/v) acetonitrile and 0.1% (v/v) TFA, and subjected to mass spectrometry (MS) analysis.

A matrix-assisted laser desorption ionization (MALDI) Orthogonal Time of Flight (TOF) Mass Spectrometer (prOTOF 2000, PerkinElmer) was used for peptide fingerprinting (MS). An in-house built MALDI interface coupled to an ion trap (Krutchinsky et al., 2001) (LCQDECAXP^{PLUS}, Finnigan) and a vacuum-MALDI ion trap (vMALDI LTQ, Finnigan) were used for amino acid sequence analysis (MS/MS). External calibration was performed using a standard peptide mixture. Computer algorithms, such as XProteo (http://www.xproteo.com) were used to correlate peptide mass fingerprint data obtained from MS analysis.

2.3.7 Immunoblot Analysis

Cell lysates and viral particles were prepared as described above. For Western blot analysis, 0.1% of the total lysate, flow through, insoluble pellet and 1% of each elution as well as the beads after elution were loaded on the gel. Samples were treated as described above. After transfer, the membrane was probed with 1 µg/mL primary antibody M2 α-FLAG mAb for 3xFLAG-tagged viral protein detection. A sheep anti-mouse horseradish peroxidase-conjugated antibody at 1:10,000 was used as the secondary antibody. Proteins were detected as described above.

2.4 Identification of Host Proteins that Interact with HIV-1 Integrase Fusions—IN Interacts With RNA Helicase A

2.4.1 Plasmid Construction

To identify cellular proteins that interact with integrase in the cytoplasm and/or the nucleus, an expression construct of integrase fused to the C-terminal of EGFP (EGFP-IN) was constructed by cloning a modified integrase (*mod integrase* gene*,* see below) into the vector, pEGFP-C1 (Clonetech). An expression vector for the fusion protein EGFP-MBP-IN(SH3) was built by inserting the maltose binding protein gene into pEGFP-C1 to make pEGFP-MBP and then cloning the C-terminal Src homology-3 (SH3) domain of integrase [residues 220-270 (Eijkelenboom et al., 1995)] at the C-terminal end of pEGFP-MBP. As controls, an IN(SH3) NLS mutant with a triple Lys to Ala mutant IN(SH3$^{K(262/3/6)A}$) (Muesing, unpublished) and an unrelated SH3 DNA binding protein, from the archaebacteria *Sulfolobus solfataricus*, Sso7d (Baumann et al., 1994), were each constructed into pEGFP-MBP. In addition, IN(SH3), IN(SH3$^{K(262/3/6)A}$) and Sso7d were inserted into the bacterial expression vector, pMAL (New England Biolabs), down-stream from the *malE* gene that encodes MBP (di Guan et al., 1988) to make constructs, pMAL-MBP-IN(SH3), MBP-IN(SH3$^{K(262/3/6)A}$) and MBP-Sso7d.

The expression vector pcDNA-FLAG-His-RHA (Zhou et al., 2003) was kindly provided by Dr. Chee-Gun Lee (University of Medicine and Dentistry of New Jersey). The FLAG-His region was removed and replaced with a triple

hemagglutinin (3xHA) tag to generate pcDNA-3xHA-RHA by PCR and molecular cloning. For immunofluorescent studies, 3xHA at the N-terminal end of RHA was replaced with 3xFLAG. An expression construct of HIV-1 integrase with 3xFLAG insertion at residue 280 (pcDNA3.1-IN-3xFLAG) was generated by first subcloning the XbaI/SacII integrase fragment of the R7/3-(XbaI/SacII)-IN-3xFLAG proviral clone into the pcDNA3.1(+) vector. The majority of the integrase gene upstream of the 3xFLAG insertion was then replaced by the corresponding part of a synthetic integrase gene (*mod*) that destroys the inherent RNA secondary structure of the native integrase gene while maintaining its amino acid sequence. It has been designed to circumvent the lack of expression of the native integrase gene and to obtain high level of IN expression – (Genbank AF422697, Muesing, 2002) and (Cherepanov et al., 2000).

2.4.2 Immunoprecipitation of IN-3xFLAG, EGPF-IN or EGPF-MBP-IN(SH_3) Fusion Proteins and Associated Cellular Factors

HEK 293T cells in a P150 plate were transfected with 15 µg of the expression construct, IN-3xFLAG, using PEI as described above. Transfection of pcDNA3.1(+) DNA was used as control. After incubation for 40 hours, cells were washed twice with PBS and lysed in a modified RIPA buffer (50 mM Tris-HCl, pH 7.4; 150 mM NaCl, 0.5% sodium deoxycholate, 1% NP-40 and 0.1% SDS) containing protease inhibitors (Roche) at a concentration of 5 mL buffer per gram of cells. After a 30-min incubation at 4 °C, lysates were sonicated with a Sonic Dismembrator 500 (Fisher Scientific) and cleared by 10 min of centrifugation at

20,000 relative centrifugal force at 4 °C. To isolate IN-3xFLAG as well as any associated protein complexes, the cleared cell lysate was incubated with α-FLAG conjugated Dynabeads at 4 °C for 1.5 hr and eluted in 30 µL 1xLDS sample buffer (Invitrogen).

pEGFP-MBP-IN(SH3), pEGFP-MBP-IN(SH3$^{K(262/3/6)A}$) and pEGFP-MBP-Sso7d were expressed and similarly lysed. The efficiency of transfection was assessed by the expression of green fluorescence. Approximately 0.9 g cells were collected and cryogenically grinded as described above. The ground powder was transferred to a 50 mL tube and 6 mL/g cells of lysis buffer of 20 mM HEPES, pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 0.1% Tween 20, 1% Triton X-100, 0.5% sodium deoxycholate, 0.5 M NaCl, 25 U/mL DNase I and protease inhibitor cocktail was added. The resulting lysate was homogenized for 10 sec with a PT 10-35 Polytron (Kinematica) and slowly rotated at 4°C for 15 min. After centrifugation for 10 min at 3600 rpm at 4°C, the supernatant fraction was used for the affinity purification experiments. Immunoaffinity purifications were performed as described (Cristea et al., 2005) using 1 hour incubation with M-270 Epoxy Dynabeads (Invitrogen) coupled with α-GFP antibodies (in-house prepared and as described above).

The MBP-IN(SH3), MBP or MBP-IN(SH3$^{K(262/3/6)A}$) fusion proteins were expressed in *E. coli* and 1 mL of crude extract in column buffer (20 mM Tris-HCl pH 7.4, 0.2 M NaCl, 10 mM DTT, 1 mM EDTA) was lysed with a Sonic

Dismembrator 500 (Fisher Scientific). After centrifugation, the supernatant was incubated with 200 µL of amylose magnetic resin according to the manufacturer's protocol (New England Biolabs). After washing three times with 1 mL column buffer, MBP fusion protein bound on amylose resin was divided into two. Each half was incubated with the cytoplasmic or nuclear extracts of CEM cells prepared using NE-PER isolation buffers according to manufacturer's instructions (Pierce). After several washes with column buffer, proteins bound to amylose beads were eluted by adding 50 µL of 10 mM maltose.

Elutions were resolved on 10% or 4-12% Bis-Tris NuPAGE gels. Each gel was fixed with 10% methanol and 7% acetic acid in water for 20 min and stained with Gel-Code Blue Stain reagent (Pierce) according to the manufacturer's staining protocol. Bands of interest in IN-3xFLAG immunoisolations were processed as described earlier and analyzed by DE-STR MALDI-TOF mass spectrometer. For EGFP-IN, EGFP-MBP-IN(SH_3) and MBP-IN(SH3) immunoaffinity isolations, entire lanes on the gel were cut into 1-2 mm pieces and protein digested with trypsin and analyzed by mass spectrometry using MALDI QqTOF MS.

2.4.3 Immunoprecipitation and Western Blots

Semi-confluent 293T cells on a 10-cm dish were transfected with 3 µg of pcDNA3.1-IN-3xFLAG and 3 µg of pcDNA-3xHA-RHA using PEI. As control, pcDNA3.1-IN-3xFLAG DNA or pcDNA3.1-3xHA-RHA was transfected with the empty vector pcDNA3.1. 36 hours after transfection, cells were lysed and

immunoprecipitated with 1 mg α-FLAG conjugated Dynabeads. Antibody-protein complexes were eluted in 1x LDS sample buffer. Immunoblotting of the elutions was performed using 1 µg/mL α-HA mAb (Covance) as primary antibody. Reciprocal immunoprecipitation was also performed using α-HA conjugated Dynabeads, and probed with 1 µg/mL M2 FLAG monoclonal antibody (Sigma).

2.4.4 Immunofluorescent Staining

For subcellular localization studies of RHA, 293T cells were plated on glycine coated coverslips in 24-well plate format and co-transfected with p3xFLAG-RHA and pEGFP-IN. After 36 hours of incubation, cells were probed for the presence of 3xFLAG-tagged RHA with 2 µg/mL M2 FLAG (Sigma) mouse monoclonal antibody. The staining and imaging using DeltaVision deconvolution microscopy was performed according to protocol described above.

2.5 Investigation of HIV-1 Integrase Post Translational Modifications

2.5.1 Characterization of HIV-1 Integrase Acetylation Mutants

The acetylation sites in HIV-1 integrase of R7/3 at lysine residues 264, 266 and 273 were mutated to arginine [K(264/266/273)R or 3R] by overlap PCR mutagenesis and verified by DNA sequence analysis. A 1xFLAG tag at the extreme C-terminus of integrase was inserted into the WT or 3R version of R7/3.

To be consistent with the original finding (Cereseto et al., 2005), the 3R mutation also was also constructed in the Bru IN background using overlap PCR mutagenesis. Viral stocks were prepared by transfecting 1 µg of each proviral DNA into 293T cells using 8 µg/mL PEI in 6-well plate format. 10 ng p24 of each viral supernatant was used to infect 1.0 x 10^6 CEM-SS cells. 18 hours after infection, cells were spun down to remove residual virus. After washing once with PBS, cells were resuspended in 8 mL fresh media. 100 µL of the infected culture was collected daily. Infection cultures were split 1:4 every 3 days. Viral titers were determined by use of an exogenous reverse transcriptase (RT) assay (Potts, 1990). The radioactive products of the RT reactions were spotted onto DE81 paper, detected by phospho-imaging and quantified using ImageQuant 5.2 software. Independent viral growth curves were generated after infection of CEM cells and monitored over a ten-day period using the accumulation of viral capsid (p24) as an indication of viral propagation. Infection in HeLa P4R5 cells was studied using 10 ng p24 of WT and 3R mutants in either R7/3 or R7/3 Bru IN. Two days post-infection, cells were processed as described above.

Other single- or double-substitution mutants at IN residues 264, 266, and/or 273 were constructed into an otherwise isogenic background (R7/3) except using the Bru IN sequence in place of the R7/3 IN. To investigate the possible co-receptor dependency of the mutant phenotypes, WT and 3R mutants in R7/3- (CXCR4 usage), and R7/3 YU-2-derivatives (CCR5 usage) were constructed by replacing the envelope coding sequence at gp120. EGFP-derivatives were constructed by

replacing the first 250 base-pairs of the *nef* region of these clones so that the successful infection could be monitored by the green fluorescence. In a 12-well plate, HeLa P4R5 cells were infected with 10 ng of EGFP-R7/3, EGFP-YU-2 and their corresponding 3R mutants to compare their dependence on CCR5 or CXCR4 entry.

A genetic assay was employed to measure the respective integration frequencies between the WT and mutant viruses. The WT and 3R mutant proviral clones were recombined in such a manner to replace the *nef* gene with a dominant selectable marker, the blastocidin D gene (bsd), which also incorporated inactivating frameshift mutations in the genes, *vpr* and *env*. Vesicular stomatitis virus (VSV)-G pseudotyped viral stocks were prepared by co-transfecting R7/3 BSD expression vectors with pCI-VSV-G into 293T cells. HeLa P4R5 cells were plated in 24-well dishes at approximately 30% confluence and infected with 5 ng p24 in 0.5 mL for 8 hrs. The medium was changed and the cells incubated for an additional 36 hrs before being trypsinized and serially diluted into medium supplemented with 5 µg/mL Blasticidin-S HCl (Invitrogen). Cells were cultured for an additional week before staining with crystal violet. Colonies were scored to determine relative integration frequencies.

For immunoprecipitations, semi-confluent 293T cells in 10-cm plate format were transfected with 3xFLAG tagged IN or co-transfected with C-terminal HA tagged p300 and pIN-3xFLAG. p300 with a 33 amino acid deletion in combination with a

catalytic point mutant in HAT domain was obtained from Dr. Richard Eckner (University of Medicine and Dentistry of New Jersey). As control, cells were either transfected with pcDNA3.1 or pHA-p300. Cell pellets were lysed in 2 mL of modified RIPA buffer (see above) with 10 mM sodium butyrate (a non-specific inhibitor of all HAT enzymes) and a protease inhibitor cocktail. α-FLAG conjugated magnetic beads were incubated at 4°C with cell extracts for 2 hours. After incubation, the immunocomplexes were analyzed by Western blotting using 1 µg/mL α-p300 (Abcam) or 1 µg/mL α-HA (Covance). Reciprocal immunoprecipitations were performed using cells co-transfected with pHA-300 and pIN-3xFLAG using either α-p300 or α-HA conjugated magnetic beads according to the protocol described here.

2.5.2 Integrase Phosphorylation by CKII in vitro

Expression plasmid pINSD-HisSol (Jenkins et al., 1995) encoding a 20 amino acid histidine (His) tag (MGSSHHHHHSSGLVPRGSH-) at the N-terminus of integrase with a single substitution of lysine for phenylalanine at position 185 (F185K) was obtained from Dr. Robert Craigie (NIH). This sequence also contains a thrombin cleavage site LVPRGSH to allow removal of His tag after purification. Proteins were expressed according to protocol (Jenkins et al., 1995). In brief, plasmids encoding the recombinant IN and site-directed IN mutants were expressed in *E. coli* BL21(DE3). Protein expression was induced by 1 mM IPTG. Cells were harvested from 100 mL of LB broth containing 100 µg/mL Ampicillin and lysed in 25 mM Hepes pH 7.5, 500 mM NaCl, 2 mM β-mercaptoethanol, 5

mM imidazole and 0.3 mg/mL lyzozyme. The samples were incubated at 4°C for 30 min, the cells were sonicated three times at 1 min intervals. After centrifugation at 14,000 rpm for 10 min, the supernatant was recovered and applied to a 0.5 x 8 cm chelating Sepharose Ni-affinity column (Pharmacia). After extensive washing with 20 mM and 60 mM imidazole in elution buffer of 25 mM Hepes pH 7.5, 0.5 M NaCl, 2 mM β-mercaptoethanol, 10% glycerol, proteins were eluted with a 1 M imidazole in elution buffer. The His tag was removed by cleavage with thrombin (Pharmacia). Optional additional purification using Heparin column chromatography (Pharmacia) was performed according to the manufacturer's protocol. Protein concentrations were measured by (BioRad). Resulting proteins were stored in 30% glycerol at –30°C. Similarly, GST fusion IN proteins were produced by replacing the His tag with GST. Proteins were purified using Glutathione Sepharose 4B Microspin Columns according to the manufacturer's directions (Pharmacia). All potential phosphorylation mutants and truncated forms of IN were constructed by restriction digestion and cloning.

The *in vitro* CKII assay was performed according to the following protocol (New England Biolabs). One µg of the purified recombinant IN proteins and various mutants were incubated with 5 U/mL CKII in CKII reaction buffer of 20 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM $MgCl_2$ in the presence of 200 µL rGTP and γ-labeled ^{32}P ATP to a final activity of 200 µCi/umol at 30°C. The reaction products were loaded onto a 4-12% NuPAGE gel and stained with Coomassie Blue and

exposed overnight at –30°C. In a titration experiment, the amount of CKII was serial diluted 1:3 starting from 50 U/mL.

Mutations of potential phosphorylation sites (S283A, T66A, T66A/S283A, T95A/S283A, S195A/S283A and T95A/S195A/S283A) were incorporated in R7/3 proviral clone by restriction digestion and molecular cloning. Viral stocks were produced and scored for syncytia formation in CEM cells as described above. Viral growth was followed for 10 days post-infection.

For the immunoprecipitation studies, N-terminal 1xFLAG tagged IN was co-transfected with Myc-CKIIβ and HA-CKIIα or α' subunits in 6 well plate format as described above. 40 hours post transfection, cells were lysed in 200 µL/well lysis buffer [50 mM Tris-HCl, pH 7.3, 150 mM NaCl, 5 mM EDTA, 0.2% Triton X-100, and CompleteTM protease inhibitors (Roche)]. After a 30 min incubation on ice, lysates were sonicated with a Sonic Dismembrator 500 and cleared by 10 min of centrifugation at 14,000 rpm at 4 °C. For co-immunoprecipitation, cleared cell lysates were incubated with the following antibodies: 1 µg/ml anti-HIV-1 integrase mouse mAb12, 1 µg/ml α-FLAG mouse mAb M2 (Sigma), 1 µg/mL α-HA or α-Myc at 4 °C for 1.5 hrs. Antibody-protein complexes were precipitated with protein A-Sepharose (Amersham Biosciences) and resolved on a 10% polyacrylamide gel. Immunoblotting was then performed as described above, using appropriate primary antibodies.

2.5.3 HIV-1 Integrase Ubiquitination (Ub)

The 3xFLAG-tag at IN residue 280 (IN-MOD) was recombined in such a way as to express either a stable and unstable integrase protein (M-IN-3xFLAG vs. F-IN-3xFLAG). These constructs specify either a methionine (M) or phenylalanine (F) residue at the N-terminus of IN, dictating the inherent intracellular stability of the produced IN protein (Mulder and Muesing, 2000). In accordance with the *N-end rule* of Varshavsky, the intracellular half-life of M-IN is greater than 24 hours while F-IN is less that 30 minutes (Mulder and Muesing, 2000). The mutant pcDNA3-HA-Ub set (5 constructs) contain four tandem repeats of HA-Ub—each mutant construct conservatively substituted with arginine in at all but one lysine residue (K6, K11, K29, K48, or K63). The converse constructs were also reconstructed. In this set, only one lysine is substituted with arginine. These constructs are designated K6R, K11R, K29R, K48R, or K63R (Nishikawa et al., 2004) and were kindly provided by Dr. Tomohiko Ohta (St. Marianna University School of Medicine, Japan).

To test IN ubiquitination, 293T cells were transfected with 5 μg of Ub-(F)IN-3xFLAG in 10-cm plate. As a control for FLAG immunoprecipitation, Ub-(F)IN was transfected alone. Cells were lysed 30 hrs post transfection in RIPA buffer and immunoaffinity isolated as described above. Magnetic bead elutions were evaluated using Western blot analysis with either 1 μg/mL M2 α–FLAG MAb or a 1:1000 dilution of a polyclonal antibody.

IN the definitive experiments, HEK 293T cells were co-transfected with 0.5 µg of pUb-(M)IN-3xFLAG or pUb-(F)IN-3xFLAG with each set of the ubiquitin expression constructs in 6-well plate format. Cell lysates were prepared with a modified RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1% SDS, 0.5% deoxycholate and 1% NP-40) that included a cocktail of protease inhibitors (Roche) and incubated for 30 min at 4^{o}C. The lysates were then sonicated for three times at 10 sec intervals, and the lysates clarified by centrifugation at 14,000 rpm for 10 min at 4^{o}C. Dynal magnetic beads conjugated with M2 α–FLAG antibody were incubated with the supernatant for 1 hr at 4 °C as described earlier. After 6 washes with lysis buffer, samples were eluted in 20 µL 1xLDS sample buffer (Invitrogen). After electrophoresis, proteins were analyzed by immunoblotting using the α-FLAG monoclonal antibody.

For mass spectrometry analysis, 293T cells were transfected with the pcDNA-Ub-F-IN-3xFLAG and the Ub expression clone with single lysine at residue 29 (pcDNA3-HA-Ub29K) on P150 plates and cultured for 2 days. Pilot experiments determined that when 29K is expressed with Ub-F-IN (all of the other reactive lysines residues are substituted with arginine in this construct), mono-Ub-substituted IN was greatly enhanced. Cells were removed from the plates by scraping and washed once with PBS. Approximately 150 mg of cells were lysed and sonicated as described above. After centrifugation, the cleared lysate was immunoprecipitated with 6 mg of Dynabeads conjugated with α-FLAG monoclonal antibody and separated on SDS-PAGE gel. The gel was fractionated

and the slices corresponding to the 36 kDa (IN-3xFLAG) and 44 kDa (IN-3xFLAG-Ub) were cut from the Coomassie-stained gel and subjected to tryptic digestion. Sequence analysis of the digested fragments was performed by DE-STAR MALDI-TOF mass spectrometry (Applied Biosystems) and MS/MS analysis was performed on QSTAR XL tandem mass spectrometry with the hybrid quadrupole time of flight configuration (Applied Biosystems). Data was analyzed using Analyst QS software from the manufacturer.

Chapter 3: Generation of Replication-Competent HIV-1 that Can Accommodate a Potent Antigenic Tag

3.1 Introduction

To identify cellular proteins that interact with a particular viral protein by high affinity immunoprecipitation, the first step is the generation of replication-competent viruses with small but recognizable tags. Potent epitopes such as FLAG or 3xFLAG provide high-affinity and specific binding to antibody and have been very popular for immunoprecipitation and protein-protein interaction studies (Prickett et al., 1989). Therefore, the 3xFLAG tag was chosen as the primary tag for our studies. As described above, replication competent viruses were recovered through a stringent biological selection from libraries of viruses that had randomly incorporated a small insertion (PmeI) at a unique location within a specific viral coding region. These 'linker-scanned' insertional libraries were created using a Tn7-based transposon system that produces a unique PmeI restriction site insertion (the 'linker') placed randomly throughout the target DNA (Biery et al., 2000). This transposon-based mutagenic strategy was adopted because of its ability above all other systems to provide random insertion into the target DNA and to as well rapidly generate many independent clones. Linker scanning mutagenesis is a genetic method that has been used to study protein structure in efforts to identify and analyze functional domains or genes (Barany,

1985; Goff and Prasad, 1991; Stone et al., 1984). It has been previously been used rapidly identify essential and nonessential in herpes viral genes (Brune et al., 1999), the *env* gene of moloney murine leukemia virus (Rothenberg et al., 2001) and the HIV-1 genome by genetic footprinting (Laurent et al., 2000). Recently, viable Hepatitis C viral (HCV) clones have been generated using a random linker insertion mutagenic approach into insert the gene for the green fluorescent protein (GFP) into the HCV nonstructural protein 5A (Moradpour et al., 2004) as well as the Sindbis (SINV) E2. In the latter case, an insertion of a 19 residue transposon or GFP was recovered (Atasheva et al., 2007; Navaratnarajah and Kuhn, 2007). Indeed, these studies have led to the characterization of viruses that are capable of accommodating an exogenous insertion and in these cases could be used for viral protein visualization or their facile purification.

Independently, we have created transposon-based linker insertion libraries in three of the nine HIV-1 viral genes: integrase (*in*), envelope gp120 V1-V3 region (herein *env*) and the viral infectivity factor (*vif*). IN is an essential, multi-functional viral protein and has been our main research interest. Vif function was unknown at the time when this study was initiated. The Env V1 to V3 loop is highly variable and we suspected that it would be likely to tolerate insertions. These transposon libraries and the tagged-yet-infectious HIV-1 particles derived from them are the first examples of applying this strategy to HIV-1. A TnsABC transposase-based system was adopted to allow random and unique insertion of a Tn7 based

transposon, a transposable element, into a plasmid of interest *in vitro* (Biery et al., 2000; Craig, 1996; Stellwagen and Craig, 1997). Importantly, Tn7 is very nonspecific with respect to site of transposition and is almost completely random with respect to insertion within AT-rich sequences, such as the HIV-1 genome. In addition, immunity is established after the completion of the first transpositional reaction ensuring only one insertional event into all but less than 1% of the target DNA molecules.

Briefly, a library of plasmids with random insertions was selected based on a drug resistant marker gene encoded by the transposed DNA segment. Subsequently, the transposon, flanked by restriction sites (PmeI) and absent in the original plasmid, was excised by PmeI digestion. This results in a library of plasmids with insertion of a 15 base-pair (bp) segment specifying a unique and blunt-ended PmeI site randomly throughout the targeted viral gene DNA. Using this methodology, we have constructed libraries of tagged HIV-1 proviral DNA each containing a unique PmeI site within HIV-1 *integrase* (*in),* its C-terminal domain (*in-ctd*)*, env* and *vif* genes.

Infection of susceptible T-cell lines with a population of viruses prepared from the PmeI insert-bearing library provides a stringent step of biological selection for viruses capable of both robust survival in culture and capable of accommodating a five amino acid foreign peptide segment (the 15 bp insertion that includes the unique PmeI site in the sequence of its cognate DNA). To recover viral DNA of

replication-competent clones, small circular viral DNA was isolated and amplified from infected cells after selection in culture over several rounds of viral replication. For all retroviruses, proviral DNA must be integrated within the host chromosomal DNA in order to serve as an active template for the expression of retroviral genome. However, typical of all retroviruses including HIV-1 is the appearance of extrachromosomal, covalently closed circular viral DNA (vCCC DNA) in the nucleus of the infected host (Gianni et al., 1975; Roth et al., 1990). Their structures are consistent with their formation by either their homologous recombination across the LTRs of a linear pre-proviral DNA precursor, the 1-LTR circle, or by host-mediated ligation of the LTRs of the linear precursor (2-LTR circle). Although the circular viral DNAs are transcriptionally silent remnants of failed attempts during retroviral integration, they are exact but circular copies of the linear integrated provirus (Jurriaans et al., 1992). The circular viral DNA is stable and persists in the nucleus for the lifetime of the infected cell (Cunningham and Muesing, unpublished results). This feature allows for the recovery of the extrachromosomal DNAs from infected cells using standard isolation procedures (Hirt, 1967). The low-molecular weight DNA obtained by Hirt extraction is an enriched source of viral cDNA. We have been able to recover 1- and 2-LTR viral DNAs prepared from infected mammalian cell cultures and then to subsequently PCR amplify the viral coding region containing specific PmeI insertions from replication-competent viral mutants.

In the end, we used transposon-mediated mutagenesis to identify sites in the

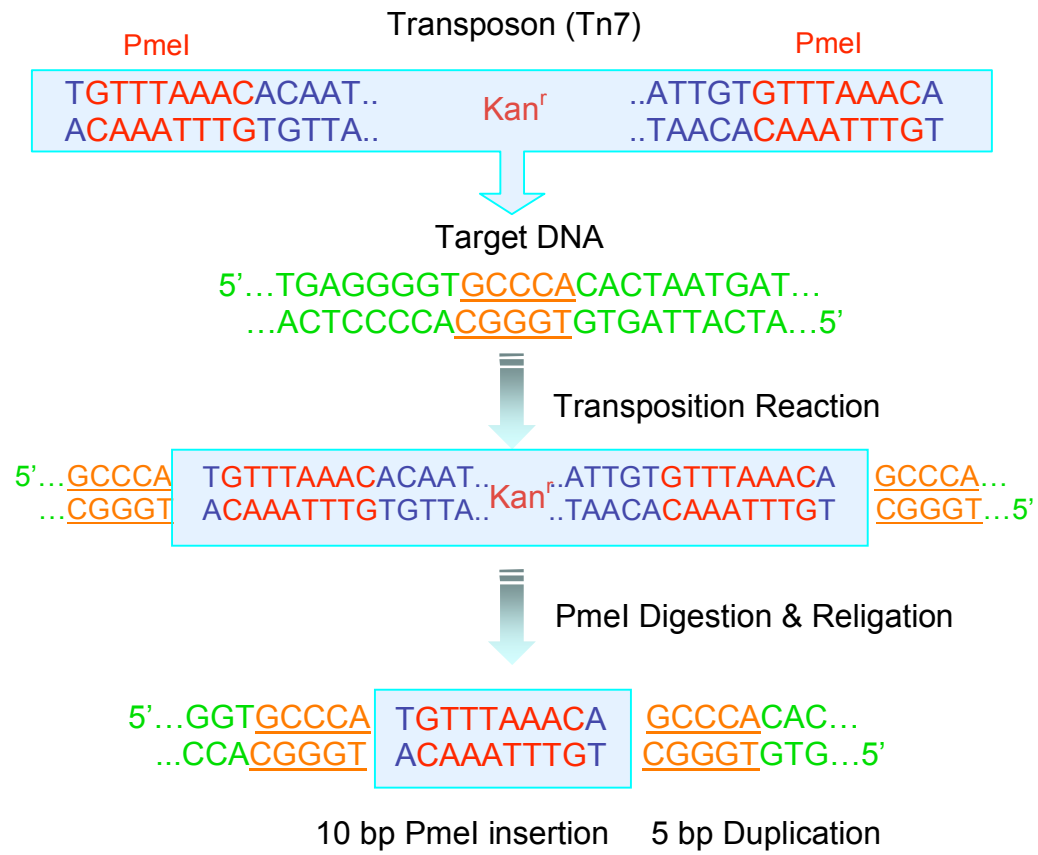
HIV-1 *pol(in), env* and *vif* genes where a 15 bp insertion could be incorporated with minimal effect on viral growth. At the sites of PmeI insertion, a larger tag such as 3xFLAG was subsequently inserted. Viable viruses harboring a 3xFLAG in the C-terminal domain of integrase (CTD), the base of the V2 loop of gp120 and in a unique internal region of the Vif protein were all recovered. Stable maintenance and recovery of these tagged viruses have been accomplished in human T-cells.

3.2 Results

3.2.1 Generation of Proviral Plasmid Libraries with PmeI Insertion

Libraries of proviral DNA with random transposon-generated insertions have been created *in vitro*. Since the transposed DNA segment (transprimer) encodes a drug resistant marker (specifying either kanamycin or chloroamphenicol resistance), recovery of plasmids with an insertion is accomplished by bacterial growth in the appropriate antibiotic. Because the transposon is unable to replicate on its own ($ori^{negative}$), only those target DNA molecules containing a transprimer insertion are recovered. Most of the transposon can be excised to leave behind a 15 bp segment specifying a unique and blunt-ended PmeI restriction site placed randomly throughout the target DNA. In four out of six reading frames, insertion leads to the addition of five foreign codons, the precise sequence depending on the identity of the target site residues duplicated upon

transposition (Figure 3.1), thereby ensuring a high probability for productive in-frame insertions.



Reading Frames

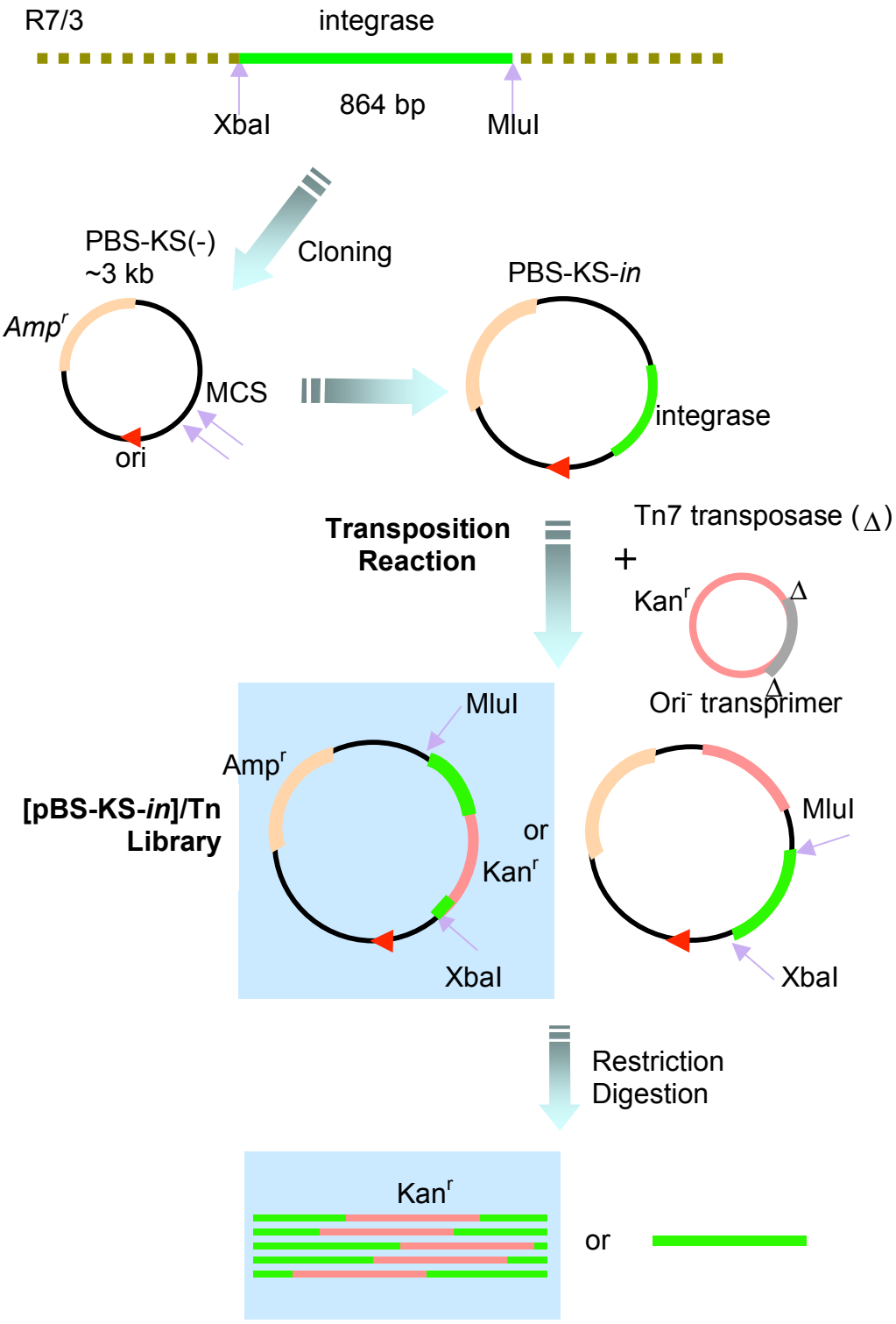
+ strand	- strand
TGT TTA AAC AGC CCA Cys Leu Asn Ser Pro	TGT TTA AAC ATG GGC Cys Leu Asn Thr Gly
CAT GTT TAA ACA GCC His Val STOP	GCT GTT TAA ACA TGG Ala Val STOP
ATG TTT AAA CAG CCC Met Phe Lys Aln Pro	CTG TTT AAA CAT GGG Leu Phe Lys His Asn

Figure 3.1 Generation of 5 amino acid insertions by a transpositional methodology, (modified from Biery, Stewart et al. 2000). For illustrative purposes, the target DNA sequence shown was selected randomly from the DNA sequence of the HIV-1 integrase.

There were several generations of libraries produced along the way for each viral gene of interest. The viral gene fragments (*in*, *in-ctd, env gp120 v1-v3* and *vif*) from the transpositional libraries were first subcloned into pBS-KS(-), a much smaller plasmid (3 kb) than the original proviral clone (16.4 kb) to maximize the number of insertional events occurring within the selected viral coding regions. The first set of libraries pBS-KS-*in* (or *in-ctd, env, vif*)/Tn was generated after *in vitro* transposition and resulted in the random insertion of the transposon carrying a Kanamycin resistant (Kan^r) gene throughout the pBS-KS plasmid. To identify those plasmids with insertions specifically in the viral coding region, a second generation of libraries, R7/3-*in* (or *in-ctd*, *env*, *vif*)/Tn was created after re-introducing each transposed viral gene into a proviral vector (R7/3). This was done by restriction digestion to release a population of fragments with random Kan^r transposon sites within the particular viral gene of interest generated in the first set of libraries. Since the fragments harboring a Kanamycin resistant transposon are 1.7 kb larger than the ones without insertions, they can be separated and purified by their mobility difference from the wild type viral gene by cleaving with an appropriate pair of restriction enzymes followed by agarose gel electrophoresis. The viral DNA fragments containing transpositional insertions were then cloned into proviral R7/3 clone to create a proviral library with random transposon insertions within the target gene. Plasmids containing the transposon were selected by plating on LB plates containing both Kanamycin and Ampicillin to ensure that the new R7/3 library contained an insertion within the targeted gene. Finally, PmeI digestion of the second transposon library eliminated the

Kan^r gene and generated the third libraries, R7/3-*in* (or *env*, *vif*)/PmeI, which contain a random 15 bp insertion within each of the viral genes originally targeted. Libraries with random 15 bp insertions were recovered on LB Amp^r plates. The scheme for generation of proviral libraries with random15 bp insertion is shown in Figure 3.2.

The number of colonies (each arising from an independent transpositional event) was determined for the second and third set of libraries. They ranged from $4x10^4$ for the *env* v1-v3 library to $7x10^4$ for *vif*. The number of clones required to assure insertion after every base pair depends on the size of the target, the desired interval between linker insertion sites and possible inclusion of the regions of the plasmid required for its maintenance (i.e. the origin of replication and the selectable marker) in the target. The number of clones that mathematically guarantees insertion after every base pair along the target viral gene was calculated based on the Poisson distribution (Pfeiffer and Schum, 1973). Here, according to the Poisson distribution calculation (Website: http://hyperphysics.phy-astr.gsu.edu/hbase/math/poifcn.html), each library contained enough diversity to assure 99% probabilistic insertion of the tag after every base in each of the targeted viral gene.



(Continued on next page)

(Continued from previous page)

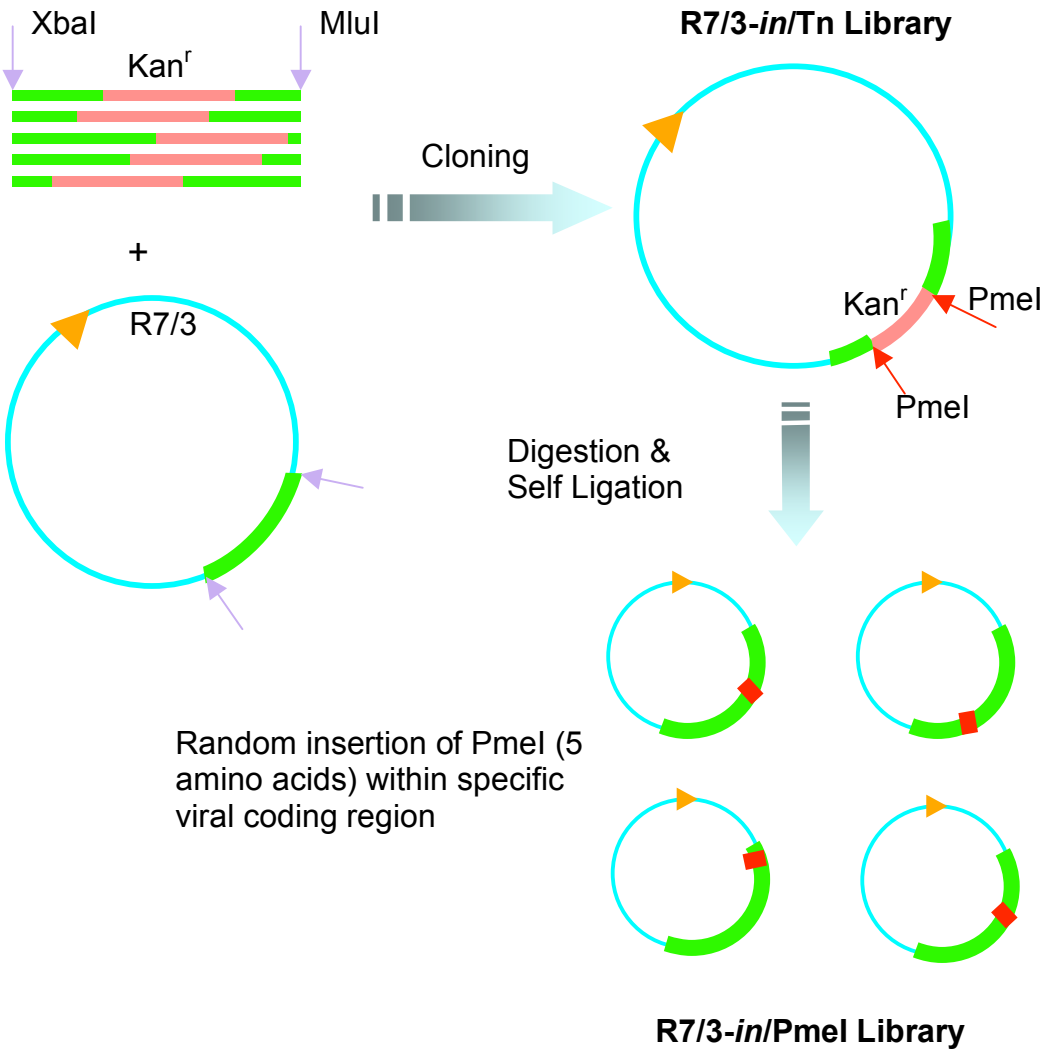


Figure 3.2 Scheme used to generate a proviral library with random and unique PmeI restriction site insertion. For illustration, the outline of the strategy employed to generate the proviral library with 15 bp PmeI insertions in the HIV-1 integrase gene.

Restriction digestion was employed to give a qualitative assessment of the randomness and uniqueness of the libraries with 15 bp PmeI insertion. DNA from the PmeI proviral plasmid library was cleaved with PmeI and the unique, stationary enzyme, BamHI. BamHI is unique within R7/3 and its restriction site lands in a non-targeted portion of the clone. Using the mixed population of transposed Pme I sites within the XbaI-*in*-MluI library as an example (illustrated in Figure 3.3A), if the PmeI insertion is unique and random within the integrase coding region, digestion will result in bands with sizes ranging from BamHI-MluI (3.3 kb) to BamHI-XbaI (4.2 kb). A visual representation of the large diversity generated in the PmeI-modified libraries is shown in Figure 3.3B. The smears on agarose gel were observed in R7/3-*in*/PmeI and R7/3-*vif*/PmeI, as well as the other proviral libraries R7/3-*in*-*ctd*/PmeI and R7/3-*env*(*v1*-*v3*)/PmeI. At least at this level of resolution, this result indicated the unique and random distribution of the transposed PmeI insertion throughout the entire targeted viral gene segment in the tagged proviral libraries.

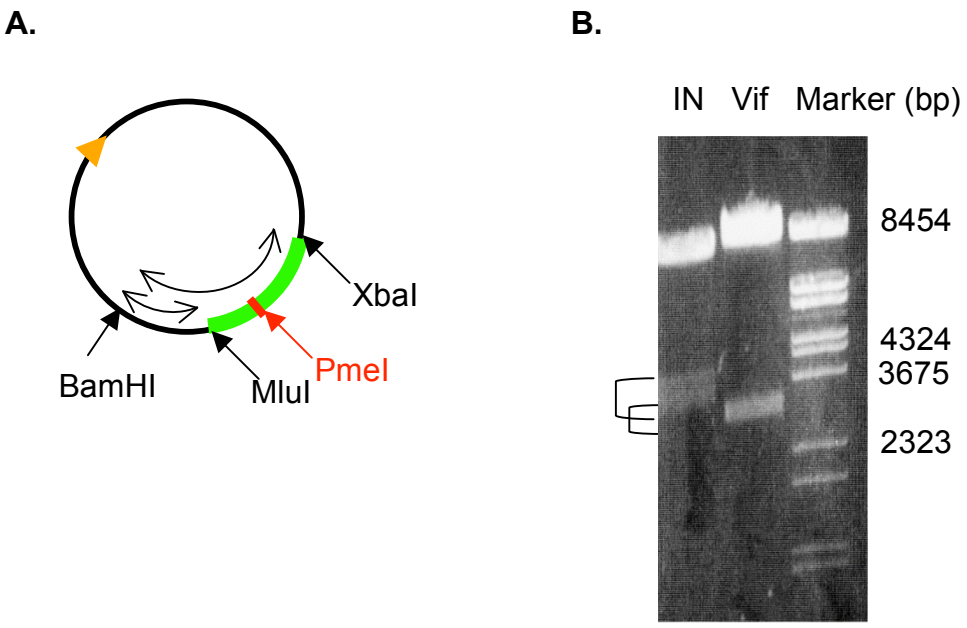


Figure 3.3 Illustration of the random distribution of the PmeI site throughout the targeted viral gene as determined by digestion with PmeI and a unique stationary site (BamHI) outside of the targeted region. (**A**) A schematic representation of the restriction digestion. Digestion is predicted to result in bands with sizes ranging from 3.3 kb (BamHI-MluI) to 4.2 kb (BamHI-XbaI) for the R7/3-*in*/PmeI library, and 2.9 kb to 3.3 kb for R7/3-*vif*/PmeI library. (**B**) Distribution of Pme I sites within the R7/3-*in*/PmeI or R7/3-*vif*/PmeI library. DNAs from the R7/3-*in*/PmeI and R7/3-*vif*/PmeI libraries were cleaved with PmeI and BamHI. Brackets indicate range of the smears.

3.2.2 Stringent Biological Selection for Replication-Competent HIV-1 with PmeI Insertion

As outlined in Figure 3.4, a population of HIV-1 particles with a 5 amino acid insertion at random locations within a specific target viral gene were produced by DNA transfection from each of the proviral plasmid libraries, the DNA pooled together from each library before transfection. Given that each virus is distinct from the other members in the library by the location of the insertion within the specific viral protein, insertions that are in incorrect reading frames or disturb critical positions in the viral proteins will not be able to pass through the next step of biological selection (viral replication). Thus, infection of susceptible T-cell lines was initiated with the heterogeneous viral stocks prepared from each of the PmeI insert-bearing libraries and followed over several days. Sustained viral propagation in tissue culture was visualized microscopically by the formation of multinucleated syncytia, which is created by the fusion of infected T cells with others via gp120 and CD4 interactions between cells in culture. A population of viruses that could survive this stringent biological selection was subsequently established during multiple rounds of viral passage. Diverse circular viral DNAs (1-LTR or 2-LTR circles) generated during this selective process were then recovered by Hirt extraction. The targeted viral DNAs within the recovered circular viral DNAs were amplified by PCR and characterized with respect to tag location in the viral gene by either restriction enzyme digestion and/or DNA sequencing.

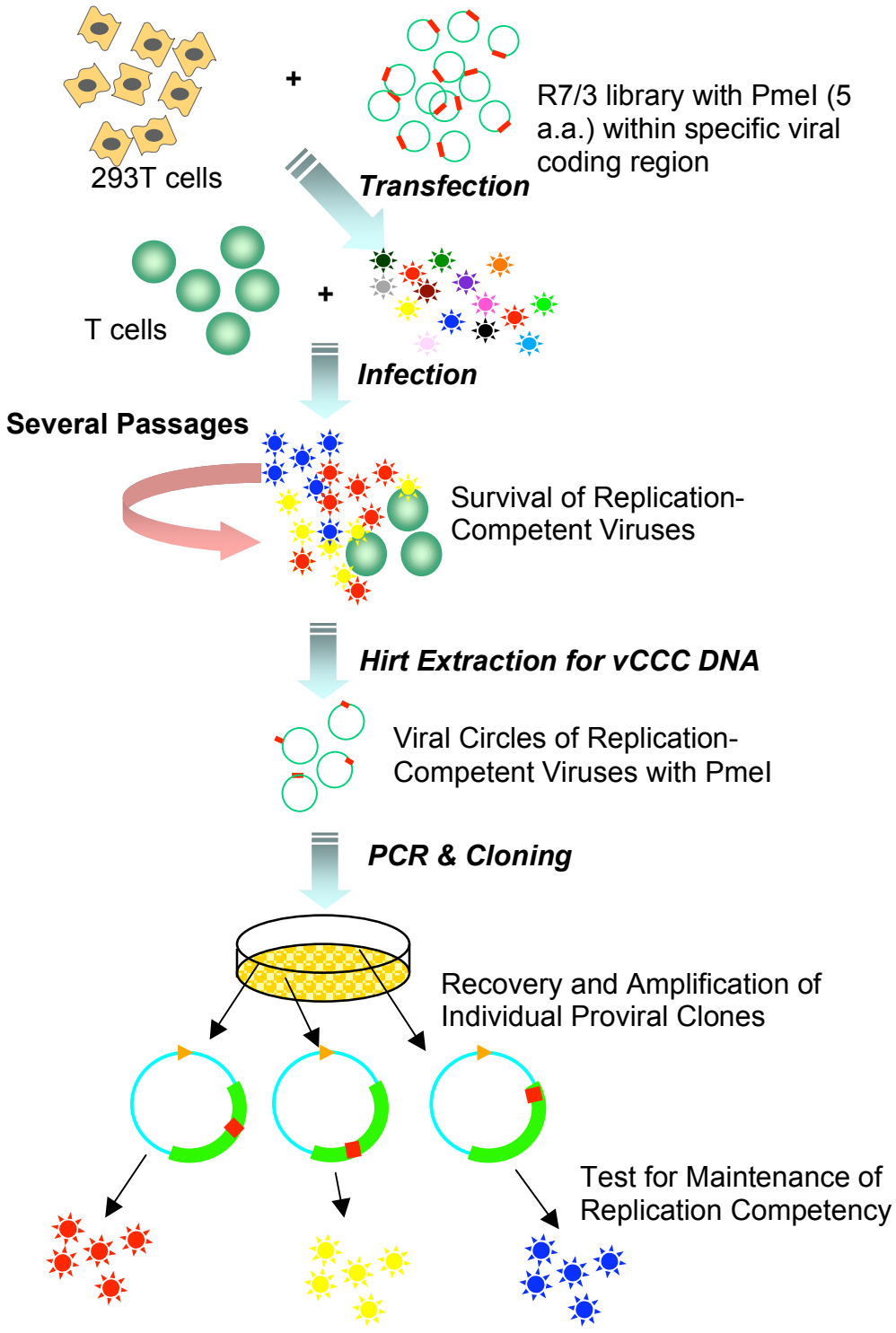


Figure 3.4 Scheme for the selection and recovery of insert-bearing replication competent viruses.

In this way, viruses that were able to accommodate insertions within the integrase C-terminal domain, Vif, and Envelope V1 to V3 region have been recovered. DNA fragments containing novel PmeI insertions were amplified by PCR of the covalently closed circular (CCC) viral DNA using oligonucleotide primers flanking the viral gene of interest. Figure 3.5 shows the PCR analysis of viral DNA recovered from viruses with PmeI insertion in the C-terminal domain of integrase (*in-ctd*). If there is insertion within this region, PmeI digestion of the PCR product will result in two fragments. Alternatively, if an intact fragment is observed after PmeI cleavage, this indicates that the PmeI insertion was eliminated. The significant amount of the intact, uncleaved PCR fragment indicates that intragenic recombination had occurred within the *in-ctd* in infected culture. Similar results were observed for other proviral libraries. As shown in Figure 3.5, the wild-type virus emerged fairly early during infection of R7/3-*in-ctd*/PmeI library. Although the wild-type virus persists, it does not dominate the infection. Interestingly, some clonally marked viral species were detected five days after infection and were able to out-compete with the wild type recombinant for the remainder of the culture period. This 'battle' was evident as demonstrated by PmeI digestion of PCR products obtained from viral CCC DNA at various times after infection. The increasing intensity of the bands corresponding to the PmeI inserted clone indicates the ability of some of the PmeI-inserted recombinant viruses to strongly compete against the wild-type over time.

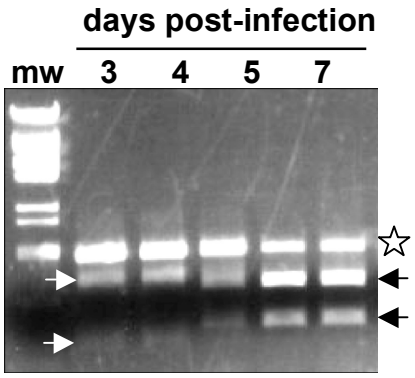


Figure 3.5 Time course of the appearance of competing viral species during selection for replication. vCCC DNA was recovered at various times after infection of CEM cells with the R7/3-*in-ctd*/PmeI proviral library (dpi, days post-infection). Two PmeI digested PCR fragments are expected for viruses harboring a unique PmeI site. An early PmeI-viral clone that appears at days 3 and 4 (white arrows) is replaced with another viral clone after day 5 (black arrows). The single uncleaved product is observed for the wild-type recombinant virus (star) and emerges and persists after day 3.

The purified PCR fragments from both 2 days- and 7 days-post-infection were cloned into the R7/3 proviral vector and individual colonies selected for DNA purification. The existence of a PmeI site was verified by restriction digestion in the individual clones. For samples collected 2 days post infection, 10-40% of the colonies examined had PmeI insertion. However, by day 7, most of the colonies isolated were wild type except for the PmeI-bearing clone from the R7/3-*in-ctd*/PmeI library. In order to determine the exact location of the PmeI insertions, all R7/3 plasmids with PmeI insertions were sequenced with primers flanking the specific coding region. For each viral gene in the study, approximately 10 clones were sequenced. Many clones selected from one library were observed to have identical or nearby points of insertion within the gene segment.

After reconstruction of each insertion into proviral clones, homogenous viral stocks were produced by DNA transfection and subsequently used to infect human T-cells in culture to confirm their replication competency. In this way, two proviral clones with the PmeI site within the *in-ctd* were studied further: an in-frame insertion at residue 280 and another added at residue 281 that resulted in translational termination of the integrase reading frame at this position. For *env:* two in-frame clones flanking the gp120 V2 loop region at residues 140 and 191, and for *vif*: two clones with in-frame PmeI insertions at residues 150 and 161. All of these PmeI viruses were confirmed as replication-competent by progressive infection over time as observed by the assessment of syncytia formation. Visual presentations of the locations of the insertions are shown in Figure 3.6.

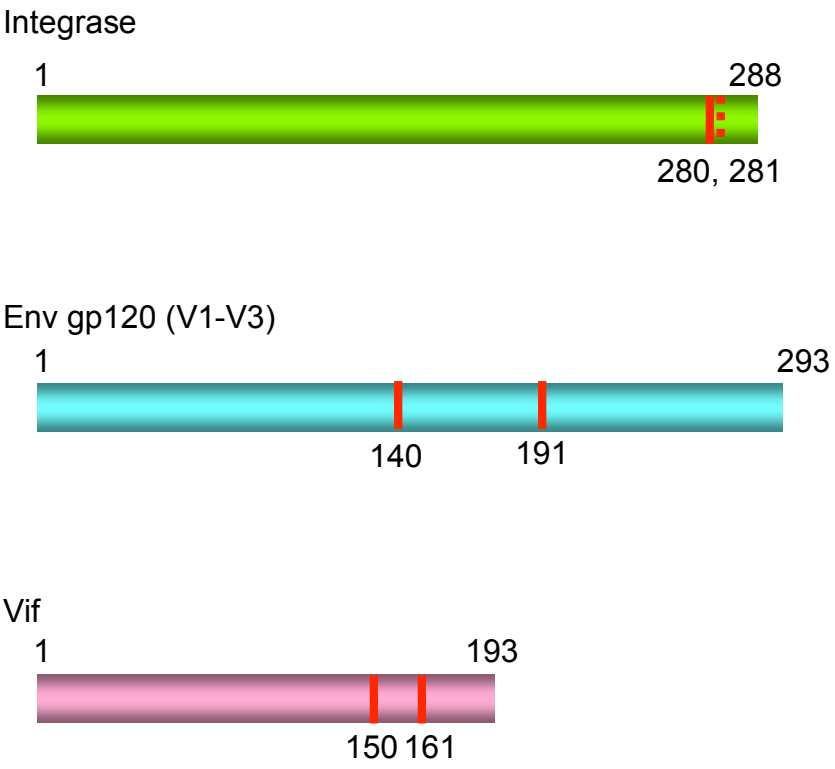


Figure 3.6 Schematic representation of the location of the PmeI site and its corresponding 5 amino acid insertion (solid lines) or the one translational stop codon insertion (dash line) within the proviral DNA. All are replication-competent.

3.2.3 Insertion of 3xFLAG Tags at the Various PmeI Sites Result in Replication-Competent Viral Clones

Each of the viable proviral clones originally selected through biological competition was linearized by PmeI restriction digestion. The linear DNA was then ligated to a 93 bp oligonucleotide pair that encodes a 3xFLAG tag at the PmeI site. The 3xFLAG tag was chosen because of its relatively small size and its extremely high antigenic reactivity with its respective, commercially available, monoclonal antibodies. The 3xFLAG oligonucleotide pair was designed so that it fused a short peptide reading frame when inserted properly but contained translational stops in all of the other five reading frames. The viral-tag junctions were designed to encode flexible amino acid adapter sequences comprised of serine and glycine codons. To eliminate the possibility of incorporating multiple monomeric units of the 3xFLAG tag into proviral DNA, a unique restriction site (BlpI) was added to the 3xFLAG tag. This site does not appear within R7/3 or with any of the R7/3-PmeI libraries. Therefore, any ligation event that result in multiple additions to the ends of the vector DNA can be “cut back” with BlpI digestion to eliminate replicate units of the tag. In addition, since BlpI recognizes a non-abbreviated palindromic sequence, re-circularization of the vector can only occur to recreate the original 3xFLAG sequence, preventing the occurrence of clones that do not. Thus after ligation, the DNA was cut with BlpI, the linear proviral DNA purified, recircularized by self-ligation and proviral clones with a single 3xFLAG insertion created (Figure 3.7).

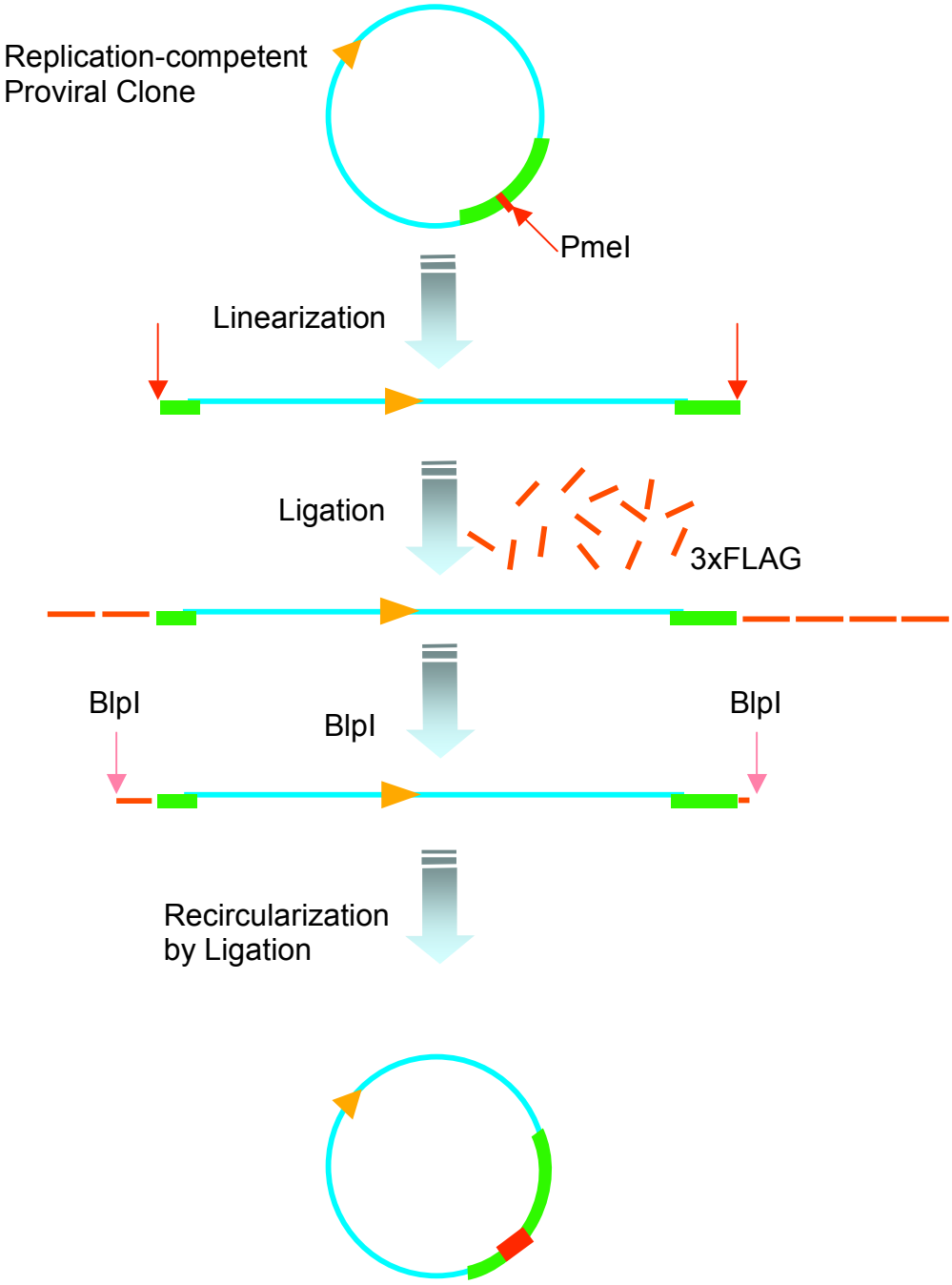


Figure 3.7 Scheme of the addition of a single 3xFLAG tag to a viable PmeI proviral clone.

R7/3 plasmids with 3xFLAG insertions were sequenced to verify the locations and sequences of the insertions. Other tags such as 1xFLAG, a biotin acceptor tag and Strep tag were also incorporated into proviral clones. Clones verified for the correct in-frame tag insertion were used to produce homogenous viral stocks and subsequently used to infect human T-cells in culture to test their respective viral replication competencies. In the end, 3xFLAG insertion at IN residue 280, Env V1-V3 191 and Vif 161 were all able to produce viable clones. They are designated R7/3-IN-3xFLAG, Env-3xFLAG and Vif-3xFLAG. The rates of syncytia formation are summarized in Table 3.

Table 3. Qualitative rates of syncytia formation in tagged viral infections.

	PmeI	3xFLAG	FLAG	SNAGG
IN (at 280)	++	+	++	+++
IN (STOP at 281)	++	NA	NA	NA
Env (at 140)	++	-	-	NA
Env (at 191)	++	++	++	+
Vif* (at 150)	++	-	+	NA
Vif* (at 161)	+++	++	++	NA
WT	+++			

* Tests performed in the Vif-restricted T-cell types, MT2 or HUT78.

3.2.4 Some 3xFLAG Insertions Produce Noninfectious Particles

As shown in Table 3, not all replication-competent PmeI viruses with 5 amino acid insertions tolerate further addition of a tag sequence. For example, no syncytia were observed when the 3xFLAG tag was inserted at residue 140 of gp120. This suggests that this 3xFLAG insertion produced a defective virus although PmeI (5 amino acid) insertion at the same site is a viable mutant. Meanwhile, to confirm that the defective virions were produced after DNA transfection, though noninfectious, virions generated by transfection of cells with the defective proviral 3xFLAG-tagged DNAs were concentrated and examined by Western blot. To investigate the extent of virion production and the fidelity of viral Gag and Gag/Pol processing, virus produced from each clone was probed for expression of the 1x or 3xFLAG tag (monoclonal M2 FLAG antibody) and probed with a highly reactive serum from an infected individual (LSS) to ascertain the overall status of the fully mature native viral epitopes. These results were compared to that of the untagged wild-type R7/3 to determine if any defects could be detected. As shown in Figure 3.8A, immunoblot analysis reveals that the Env gp120 with 3xFLAG tag at residue 140 had low levels of 3xFLAG expression, the disruption in expression probably related to the characteristics of insertion at that particular location. Probing with patient serum LSS also indicated a reduction of processed viral Gag proteins relative to the wild type virus. Similar results were observed in a proviral clone with a 3xFLAG insertion at Vif residue 151.

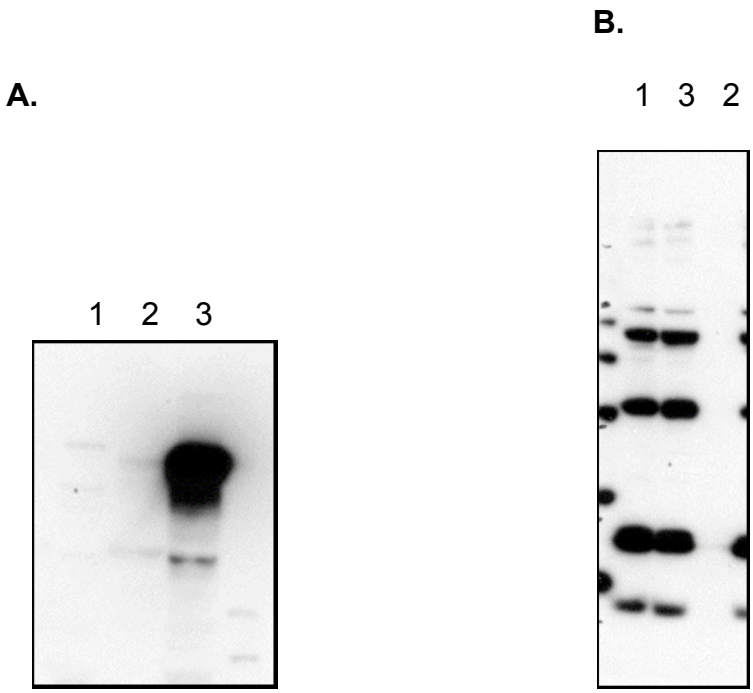


Figure 3.8 Immunoblotting of viral particles with 3xFLAG tag located within Env gp120 from producer cells. Lane 1: wild type, lane 2: 3xFLAG insertion at Env V1-V3 residue 140, lane 3: insertion at Env V1-V3 at residue 191. Probed with (**A**) M2 monoclonal FLAG antibody or (**B**) LSS.

3.2.5 The Expression of 3xFLAG Tagged Viral Protein is Stable and Maintained over the Course of Long-Term Infection

Western Blot analysis of infected cultures or precipitated viral particles of R7/3-IN-3xFLAG, Env-3xFLAG and Vif-3xFLAG collected up to 10 days post infection were performed to investigate whether the 3xFLAG fusion proteins were stably maintained during several cycles of infection. As shown in Figure 3.9, a band of the expected molecular weight of approximately 36 kDa, corresponding to the IN-3xFLAG fusion protein, was detected in viral particles concentrated from CEM cells infected with R7/3-IN-3xFLAG by both the M2 monoclonal FLAG antibody and a monoclonal integrase antibody 6G5 (Nilsen et al., 1996) that recognizes an amino-terminal epitope common to both the tagged and untagged integrase. Thus, comparison of the reactivity of the two monoclonal antibodies is a measure of the relative stability of tagged integrase over time. Meanwhile, correctly processed polyprotein of the expected molecular weight of 55 kDa and 24 kDa was detected by Western blot using the LSS human antisera. These results demonstrate that the 3xFLAG moiety is retained during infection and does not interfere with viral polyprotein processing. In addition, as shown in Figure 3.9, there was no reduction in the size of the band corresponding to IN FLAG or 3xFLAG fusion proteins 7-10 days post-infection when probed with a monoclonal antibody against native integrase. Importantly, these results indicate that the 3xFLAG fusion proteins do not revert to wild type IN and are expressed stably during progressive and extended infection. Similar findings were obtained for the Vif 3xFLAG and Env tagged viruses (Figure 3.9).

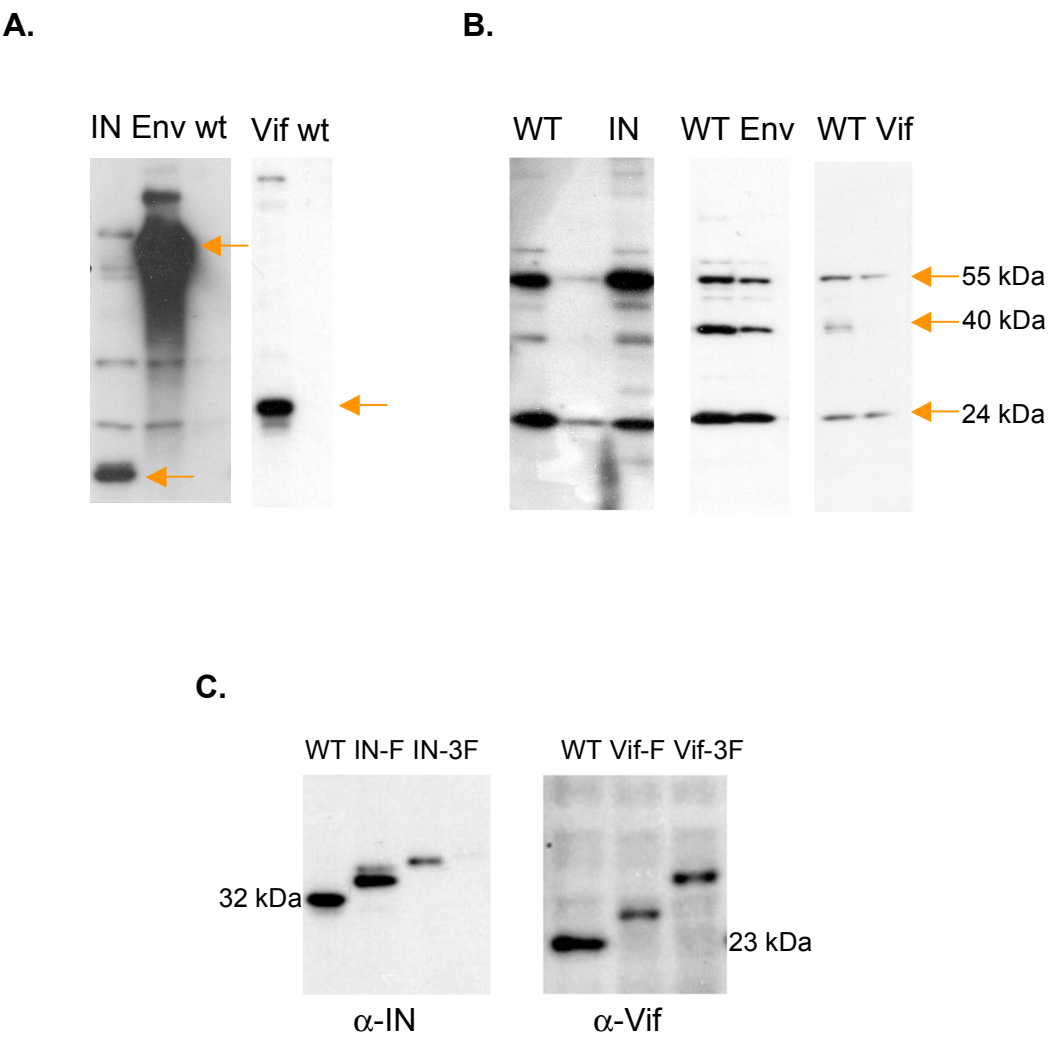


Figure 3.9 Immunoblotting of HIV-1 with 3xFLAG insertion 7 days post-infection. Viral particles were probed with either (**A**) the M2 monoclonal FLAG antibody, (**B**) human antisera LSS, or (**C**) a monoclonal antibody against integrase (6G5) or Vif (#809, NIH AIDS Research and Reference Reagent Program).

3.3 Discussion

Many studies have focused on understanding the interaction between HIV-1 and the host cell (Turlure et al., 2004; Van Maele et al., 2006). Knowledge of the different cellular partners of viral proteins would be invaluable for understanding the biology of the virus with the potential for designing interaction-specific antiviral agents. Recovery of replication-competent tagged viruses with insertions through biological selection coupled with high affinity immunoprecipitation is one approach to identify cellular proteins (and their complexes) that interact with a particular viral protein. Here I describe the generation of replication-competent HIV-1 specifying the potent 3xFLAG tag insertion in the coding regions of the integrase (IN), envelope (Env) gp120 V1-V3 region and Vif proteins. An essential component to this effort is the identification of a site or sites within the targeted viral gene that can accommodate a small foreign insert. Toward this goal, we used Tn7-mediated *in vitro* transposition to create a pool of random insertions within each of the targeted viral coding regions. Each member in the library differs from the others in the location of the insertion. Only proviral clones with insertions that do not significantly disrupt the function of the essential viral protein can complete multiple replication cycles in culture—a most stringent step of biological selection. Permissive sites (within the IN-CTD, the gp120 V2 loop and an internal region within Vif were identified using this approach, only to be subsequently utilized as a blueprint for the insertion of larger peptide elements (i.e., 3xFLAG) that could be accommodated without loss of replication potential. Most of the epitope tagged viruses that were generated in this fashion are robust

for viral spread or moderately impaired for replication. The IN (32 kDa) and Vif (23 kDa) proteins are increased in size by more than 10% with the addition of the 3xFLAG tag (3.3 kDa). Indeed, it is quite remarkable that even small insertions are tolerated, given that each viral protein must be required to act in concert with both other viral proteins and a potential multitude of host factors to affect their many independent functions during cycles of productive infection. One aspect of the tag design that may facilitate the 3xFLAG being tolerated internally is the fact that the tags have flexible adaptor Gly-Ser adaptor amino acids flanking both ends of the peptide; in a position, perhaps, to be less likely to disrupt the structure of surrounding domains. Regardless of the reason(s), these experiments indicate that even given the constraints that are uniquely imposed on the primary, secondary, tertiary and quaternary structure of the “multi-tasking” viral proteins, some flexibility must remain for the incorporation of insertion without overt loss of protein function(s).

The goal of these studies is to obtain identity profiles of host protein complexes that interact with specific viral proteins at specific times during the HIV-1 life cycle. To this end, we have used a genetic approach to select viruses that can stably incorporate an efficient epitope tag within an essential viral protein yet remain capable of robust replication competency. The random insertion approach has been very important to our success, especially valuable to those viral proteins whose structure and function are poorly understood. Instead of a hypothesis-driven-protocol used to predict locations within the viral protein most

amenable for tagging, we have employed a powerful biological selective tool (viability) that does not rely on preconceived notions of tag placement. Thus, recovery of replication competent viruses that are able to tolerate an insertion within essential viral gene products is without bias. Using this approach, this strategy has allowed us to obtain potent antigenic tags in infectious viral particles in an effort to define HIV-1-host interactions. To date, we have created transposon-based tagged libraries in three essential viral gene products: integrase, envelope and the viral infectivity factor, Vif.

Transposon-based mutagenesis coupled with a selection based on viability is a rapid way to ascertain those regions in a viral proteome that be modified without cost to propagation. Here, based on the complexity of the libraries obtained, the mathematical prediction is that transposition occurs at every base pair within the target gene segment. Ideally, Tn7 transpositional mutagenesis is random. However, the possibility still exists for some regions to exhibit a higher incidence of insertion. Indeed, Tn7 transposase has an unbiased affinity for AT-rich regions but exhibits a low efficiency of transposition into more GC-rich sequences (Kumar et al., 2004). Evidence of target skewing of the Tn7 transposase is shown in Figure 3.3. Here, signals corresponding to uniform-sized fragments are detectable through the background smear of random insertional events.

Many viral proteins have multiple roles in HIV-1 life cycle. For example, HIV-1 IN, originally defined by its role in viral DNA integration (Van Maele et al., 2006), has

been implicated in the efficiency of reverse transcription (Wu et al., 1999), viral nuclear import (Hearps and Jans, 2006) and DNA repair (Mulder et al., 2002). In addition, at least the enzymatic features ascribed to IN seems to rely on the ability of the protein to adopt various multimeric configurations. Therefore, it is not expected that functional integrase can tolerate any appreciable insertional modification (especially the size of a 30 amino acid peptide, 3xFLAG tag). Although fusion of an eight amino acid segment (1xFLAG) to the C-terminus of IN has been explored previously (Petit et al., 1999), a larger epitope (3xFLAG) within the body of IN has not been described in the context of a replication competent viral clone. The proximity of our 3xFLAG tag to the C-terminus of IN (at the 280th residue of the 288 amino acid protein) is in accordance with both structural and antibody data of the native protein that support the notion that residues 270-288 are highly flexible, certain conditions favorable for the adoption of an alternate configuration (Asante-Appiah and Skalka, 1997). Failure of both NMR and X-ray studies to provide structural details of this region is also consistent with this region exhibiting dynamic properties (Chen et al., 2000). If pushed further, we believe that even more infectious clones can be recovered from our saturated insertional library of the IN 270-288 region. Interestingly, although a viable C-terminal 1xFLAG proviral clone has been previously obtained (Arhel et al., 2006; Cereseto et al., 2005; Petit et al., 1999), there have been no reports of successfully using this particular virus for the identification of cellular proteins interacting with IN during infection. In the context of the complexes that IN encounters during its infectious cycle through the cell (i.e., RTC, PIC, nuclear),

it is possible that the three-dimensional space occupied the relatively concise 1xFLAG-tag is not readily accessible to antibody under conditions that favor the recovery of its potential binding partners. To support this notion, our own immunoisolation experiments using virus with 1xFLAG insertion in integrase have shown that the smaller 1xFLAG tag was not efficient for immunoprecipitation.

Envelope gp120 is remarkably diverse (Wyatt et al., 1998) and is by far the least conserved protein of HIV-1. However, HIV-1 Env must conserve determinants that mediate CD4/co-receptor binding (Zhou et al., 2007) and although substantial structural rearrangements of the gp120 glycoprotein occurs upon receptor binding, fusion and entry of the virus into the cell is rarely subject to antibody-mediated neutralization (Rits-Volloch et al., 2006). In our study, the V1 to V3 variable loop region (Env amino acids 126-338) was initially chosen for targeting (amino acids 62-345, see Chapter 2) because of its diversity in sequence (Siwka et al., 1994), including amino acid deletions and insertions and the most likely of all viral proteins to tolerate modification. Indeed, many escape mutants to neutralizing antibodies against HIV-1 have been mapped to the V2 region (Shibata et al., 2007). Interestingly, both 5 amino acid insertions identified in our initial random screen for viable PmeI-tagged viral mutants are located on opposite sides of the base of the V2 loop (at residues 140 and 191). However, when 3xFLAG was subsequently inserted into each of the clones, one of the two viral clones (insertion at residue 140) was replication defective and resulted in dramatically reduced 3xFLAG tagged Env expression (Figure 3.8). The

replication-defective phenotype of this ENV-3xFLAG tagged mutant was independently verified by the lack of observable syncytia or cell death in T-cells over time in culture (Table 3). There are several possible reasons for the intolerance of the 3xFLAG tag at position 140. The placement of a relatively large acidic epitope tag might disrupt proper protein folding and stability. It is also possible that steric hindrance has played a role in the intolerance of the 3xFLAG tag at this particular site of insertion. For example, insertions into the V2 loop might disrupt the overall flexibility of Env, resulting in inability for a predicted conformational shift during receptor binding. Also, the V2 140 mutant virus might not have survived concentration of virions during centrifugation, or perhaps the number of defective virions produced was significantly less than the wild type. This result indicates that the biochemical properties of identical amino acids inserted into seemingly structurally related positions may have completely different consequences for a given protein, further insertion tolerated at one location but not another.

The interaction between HIV-1 Vif (192 amino acids) and cellular factors have been extensively studied in recent years (Ehrlich and Yu, 2006; Schrofelbauer et al., 2004; Sheehy et al., 2002; Xiao et al., 2006). Two PmeI insertion mutants survived selection in Vif-restricted cells, one at residue 150, another at position 161. Significantly, the PmeI clone at position 161 was recovered from selections from two independent transposed Vif libraries. Insertion of the 3xFLAG at this site produced a replication competent virus capable of robust propagation in Vif-

restricted cell types. Interestingly, this insertion is located near a Vif zinc binding domain documented to be involved in direct binding of the Cul5 E3 ubiquitin ligase and required for Vif activity (Xiao et al., 2006). In addition, point mutational analysis has shown that both Leu 163 and Leu 169 are critical for the functional integrity of Vif (Yu et al., 2004). The unique location of the Vif-161-3xFLAG clone identified here might confirm and reveal additional information about Vif-cellular protein interactions. An infectious HIV-1 clone in which the end of Vif carried a hemagglutinin (HA) tag has been rationally constructed in an effort to immunoprecipitate Vif from HIV-1 infected cells. This strategy subsequently identified cellular proteins that interact with Vif (Yu et al., 2003). In this study, an antibody against HA immunoprecipitated HA-tagged Vif from the lysates of HA-tagged virus infected cells but not untagged Vif from the wild type control samples. Three cellular proteins co-precipitated with Vif-HA but were absent from the wild type control. They were identified by mass spectrometry as Cullin 5, elongins B and C. It will be interesting to see if different host proteins can be isolated during infection with viruses tagged at different locations.

Initially, a transposon insertional library was constructed for the whole HIV-1 genome to insert a FLAG tag. However, wild type virus quickly overtook the culture and the FLAG insertion was quickly eliminated during an early cycle of infection. Although Western blot analysis using α-FLAG revealed the presence of the FLAG tag in the producer cells transfected with the pooled whole genomic FLAG library, the signal disappeared at some point during subsequent rounds of

viral passage. A similar phenomenon is observed in Figure 3.5, where the PmeI site was almost entirely eliminated in the recovered viral DNA. We suspect that intravirion recombination is responsible for the elimination of tagged viral genes. Since HIV-1 has two copies of RNA as its genetic information, recombination can occur during reverse transcription when reverse transcriptase “stitches” together viral DNA from both single-stranded RNA templates, the final DNA product, is a composite of the original genomic RNAs. If the two RNA templates contain insertions at different genomic locations, a cross over event occurring at different points between the two can restore the wild type, untagged viral genome. This created a significant problem in our study. Because of the way the viral stocks were generated (DNA transfection of a heterogeneously-tagged pooled library), it is of high probability that two viral genomic RNAs, each specifying a tag in a different location, could be copackaged into the same virion. As a result, the genesis of the wild type virus could rapidly overtake a culture especially in cultures where the tagged viral clones are only moderately compromised for viral growth (i.e., the likely case when there is an exogenous insertion). In order to reduce this problem, insertional libraries were constructed in individual viral coding regions or domains ranging from 300 bp (*in-ctd*) to 800 bp (*Env v1-v3*), instead of the whole HIV-1 genome (10 kb). Since recombinational events occur with less frequency between closely spaced sequences, this strategy greatly reduces the chances for the appearance of the wild type virus.

We started with libraries in which a small insertion (5 amino acids) was first

constructed in the hope that less alteration to the natural viral protein would correspondingly produce a larger subset of tagged, replication competent mutant viruses capable of competing with the wild type virus over extended periods in culture. Moreover, the viral supernatant was serially collected over several days, monitored by PCR and PmeI digestion to ascertain the level of wild type contamination at each time point (see Chapter 2). The appropriate samples then processed for recovery of the tagged viral sequence.

There are several advantages using an exogenous tag instead of antibody raised directly against HIV-1 viral proteins. First, although numerous antibodies have been generated against HIV-1 viral proteins, for example integrase (Bizub-Bender et al., 1994; Nilsen et al., 1996), the use of these monoclonal antibodies for *in vivo* immunoprecipitation or immunocytochemistry has not been successful. Second, some of the antibodies generated exhibit inhibitory effects on the *in vitro* activities of integrase (Nilsen et al., 1996; Yi et al., 2000). It is possible that the epitopes of these antibodies are in the critical regions of the viral protein are required for functionality or for proper protein binding and, therefore might not be accessible for the specific monoclonal antibody binding during infection. A third advantage of using our approach is that sites identified within particular a particular viral protein that can tolerate the addition of an exogenous tag can be used for other applications such as real-time visualization by inserting a different tag such as tetracysteine or GFP.

Epitope tagging has become a powerful tool for the detection and purification of expressed proteins. This methodology has been used in studies such as protein-protein interaction, protein localization and immunoprecipitation. Among the most popular tag systems is the FLAG system (Sigma) that relies on the FLAG octapeptide (DYKDDDDK) fusion tag, a hydrophilic and immunogenic purification tag specifically and rationally designed for antibody-mediated identification and purification of recombinant proteins (Hopp et al., 1988). FLAG system can be used with one of three monoclonal antibodies, M1, M2 and M5, each recognizing and binding to the FLAG epitope with different specificities that depend on the position of the FLAG peptide in the fusion protein. M2 α-FLAG reacts with the fusion proteins with the internal FLAG epitope, or the N- or C-terminal FLAG tag and prefers the sequence Asp-Tyr-Lys-Xxx-Xxx-Asp-Xxx-Xxx (Slootstra et al., 1997). In a flow cytometry assay, the monovalent dissociation or affinity constant K_d of the M2/FLAG interaction was determined to be about 8.0 nM (Buranda et al., 2001). Placing multiple FLAG epitopes in tandem (i.e. the 3xFLAG tag) has been designed for intracellular expression with increased sensitivity of detection. The presence of three FLAG epitopes in series greatly increases the detection limit. In addition, the M2 antibody is also capable of reacting with an alternate epitope embedded within the 3xFLAG sequence. Thus, the 3xFLAG system allows for increased sensitivity and detection of the FLAG epitope tagging system while retaining the benefits of 1xFLAG expression. These benefits include the hydrophilic nature of the tag and its small size. Detection of 3xFLAG using the M2 α-FLAG is at least 20-200 times more sensitive than any other commonly

used tag, and 10-20 times more sensitive than the 1xFLAG tag (Hernan et al., 2000). Therefore, it was an obvious choice for our studies here.

One of the major obstacles to the visualization and characterization of HIV-1 proteins is the absence of strongly reactive antibodies to the native viral proteins. The 3xFLAG tag is an extremely potent tag and the 3xFLAG tagged viruses described here therefore represent an attractive system for proteomic studies. Such studies have been successfully initiated and will allow us to gain a dynamic view of the requisite host factors and their complexes involved in HIV replication. Moreover, the strategy we describe here will greatly facilitate the visualization of functional viral proteins in living cells at various points during viral life cycle.

We used the linker scanning mutagenesis to generate comprehensive libraries of clones each bearing a single oligonucleotide insertion at a random position in the gene of interest. The selection of replication competent clones from these libraries can also be investigated by genetic footprinting (Singh et al., 1997). This methodology has the ability to analyze in parallel a large set of molecularly defined mutations that can be subjected to a specific functional selection. This technique is generally applicable for saturation mutagenesis and high-resolution functional mapping of cloned DNA sequences. In our case, the HIV-1 1- or 2-LTR viral DNAs can be isolated from a saturated library with 15 bp insertions before and after infection of target T cells. Alternatively, viral genomic RNAs can be recovered from viral particles and the viral coding region of interest amplified by

RT-PCR. The mutations represented in each sample can be analyzed by PCR using one primer specific to a site in a mutagenically static region and a second primer corresponding to the inserted oligonucleotide (i.e., 3xFLAG). As the result, an insertion at a particular location gives rise to an electrophoretic band of discrete mobility. For each mutation, a PCR product of unique length will be generated, depending only on the position of the inserted oligonucleotide in the gene. For the whole library, this results in a ladder of bands, each representing a specific insertion. Mutants in which the inserted sequence disrupts a feature required for the viral replication would fail the selection and no band would appear at the position. The corresponding bands are therefore absent from the ladder of PCR products obtained from the library after infection, giving rise to a genetic footprint representing features of the gene in primary amino acid sequence that are essential for infection. Because the sequence of the inserted oligonucleotide is known, and its position can be inferred precisely from the electrophoretic mobility of the corresponding band, the precise location of insertions that disrupt gene function can be determined without isolating or sequencing individual clones.

In summary, the keys to the successes of our approach include the random introduction of a small insertion into a limited region of the HIV-1 genome and a step of stringent biological selection to obtain those tagged viruses that remain stoutly infectious in culture. In turn, these viruses can be readily utilized as templates for the further modification of the viral genome allowing the

introduction of a variety of powerful antigenic and biochemical tags depending on the nature of downstream applications.

Chapter 4: Characterization of Replication-Competent Tagged Viruses

4.1 Introduction

Using a transposon-based approach, we have generated HIV-1 proviral libraries with a 5 amino acid insertion randomly and uniquely placed throughout the integrase (IN), envelope gp120 v1-v3 loop (Env) and the viral infectivity factor, Vif. By selection of viable insertion expressing variants, we have identified positions within these viral proteins suitable for insertion of a much larger peptide, such as the 3xFLAG-tag, without a strong negative effect on virus replication. Viable viral clones with 3xFLAG tag insertions were used for the affinity purification of virus-host protein complexes that associate specifically with the tagged protein in HIV-1 infected cells. This strategy is sensible only if the tags do not interfere with the protein's normal activities during the viral replication cycle. Additionally, valuable structure-function information was obtained for each of the three viral proteins. Since the number of independent clones for each gene greatly exceeded the number of nucleotides for each gene segment, the location(s) of the sites amenable for insertion providing a mapping tool for exposed surfaces or subdomains within the specific viral protein. The virological and biochemical properties of these tagged clones are characterized herein. We also demonstrate that all three viruses with 3xFLAG tags can be quantitatively immunoisolated from infected cultures.

The 5 amino acid insertions and 3xFLAG insertion selected from integrase library is located near the end of the IN-CTD (residues 220-288). The IN-CTD contributes to DNA binding and the oligomerization necessary for the completion of integration (Coffin et al., 1997) and is linked to the catalytic core by residues 195-220, an extension of the final helix of the core domain. Crystal structures of the IN core and CTD shows that residues 271-288 are not clearly defined in density maps (Chen et al., 2000). This indicates the highly dynamic and disordered nature of this region precisely where the 5 amino acid or 3xFLAG tag insertions were allowed.

The gp120 variable regions are surface exposed and exhibit natural variation. The Env library was constructed to capitalize upon the highly variable V1-V3 variable loops where the insertion of an artificial tag might easily be tolerated without loss of function. In the end, a 3xFLAG insertion in the V2 loop region resulted in a replication competent viral clone. Previous studies have demonstrated that changes in the V1/V2 variable loop affect envelope glycoprotein receptor binding and virus entry (Fung et al., 1992; Kolchinsky et al., 2001). Here, we investigate the consequence of an artificial 3xFLAG insertion in the V2 region, and whether an antibody directed against the inserted sequence can neutralize the clone.

Since the virus with a foreign epitope like the 3xFLAG tag in the V2 region is replication competent, it is likely that this insertion did not result in significant effect on viral entry. Neutralizing antibodies often recognize regions of viral envelope glycoproteins that play pivotal roles in receptor binding or other aspects of viral entry. A simple neutralization assay can be employed to investigate whether or not the exogenously added epitope is accessible to its cognate antibody on the assembled Env trimer and also to probe if normal gp120 structure or function is maintained as revealed by the kinetics of infection. Herein, we investigate the neutralization of the modified mutant virus by a known neutralizing antibody against HIV-1 Env, as well as neutralization by an antibody directed against the artificial epitope tag.

We have constructed an infectious HIV-1 clone with a 3xFLAG tag inserted into the Vif region at residue 184. The vif open reading frame is present in most lentiviruses and is required for viral replication and pathogenicity *in vivo* (Simon et al., 1998). HIV-1 Vif is essential for viral evasion of host antiviral factor APOBEC3G (Sheehy et al., 2002). The ability of Vif to suppress antiviral activity of APOBEC3G is specifically dependent on Cul5-Skp1-cullin-F-Box (SCF)-like complex (Yu et al., 2003). The site for 3xFLAG insertion is in proximity to the identified binding region of the SCF machinery. Verification of the known and unknown host factors involved in Vif activities during infection may provide further insights into understanding the functions of this important viral gene product.

4.2 Results

4.2.1 Viability of HIV-1 Virions Harboring Tag Insertions in IN, Env and Vif

As described in the previous chapter, when compared with the wild-type virus by microscopic examination, the tagged viruses were as rapid or nearly as rapid for syncytia formation during infection. A panel of independent assays investigated the effect of the insertions on viral replication quantitatively. Viral growth curves were generated for R7/3-IN-3xFLAG and R7/3-Env-3xFLAG by measuring daily the concentration of the viral p24 core antigen accumulation in the culture supernatant up to 11 days post-infection. Viral loads from wild type virus were used as control. As shown in Figure 4.1A, the viral growth curve indicated that both HIV-IN-3xFLAG and HIV-Env-3xFLAG replicated almost as efficiently as the parental clone R7/3 in CEM cells, with slight delay in growth. They exhibited growth kinetics of less than one log when compared to that that of wild type R7/3. Growth curves for the Vif 3xFLAG tagged virus shows that it was as infectious as the wild type control in the Vif-restricted cell type, MT2 (Figure 4.1B). These results indicate that the relatively large tag in these viral proteins do not disrupt the normal functions of the proteins during infection of these cell types.

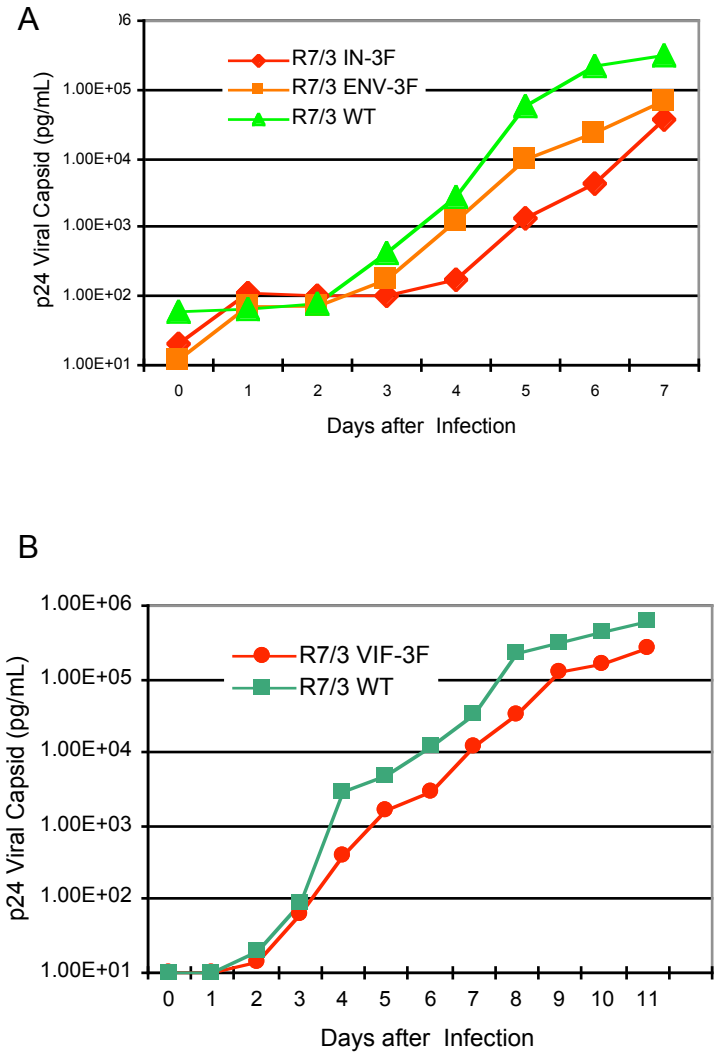


Figure 4.1 Viral growth curves of (**A**) HIV-IN-3xFLAG, HIV-Env-3xFLAG and wild type R7/3 virus in CEM cells. (**B**) R7/3-Vif-3xFLAG and wild type R7/3 virus in MT2 cells, a Vif-restricted cell type.

The growth properties of viral clones with insertions were further verified by one-step growth analysis. In a single-step assay to determine the relative infectivity, an expression vector for the vesicular stomatitis virus (VSV)-G protein was co-transfected with the parental wild type viral DNA and those corresponding to, R7/3-IN-PmeI, 1xFLAG, 3xFLAG, or the biotin-acceptor tagged viruses SNAGG. The viral constructs are also defective for HIV-1 envelope expression to allow the efficient pseudotyping with the VSV-G envelope. The relative infectivity quotient of the VSV-G pseudotyped viruses was investigated in the LuSIV (SIV LTR-Luciferase) indicator cell line. Viruses with various tags in integrase were found to be nearly infectious, or as infectious as the wild type virus, as shown in Figure 4.2. The results show that the viable mutants are capable of infection, some exhibiting levels of infection the same as the wild type. As a rule, viruses with smaller inserts (PmeI insertion or 1xFLAG insertions) are more infectious than ones specifying larger tags (3xFLAG). In addition, the one step infectivity assay for R7/3-IN-3xFLAG correlates well with the replication properties as measured in the multiple-cycle growth curves experiments described above.

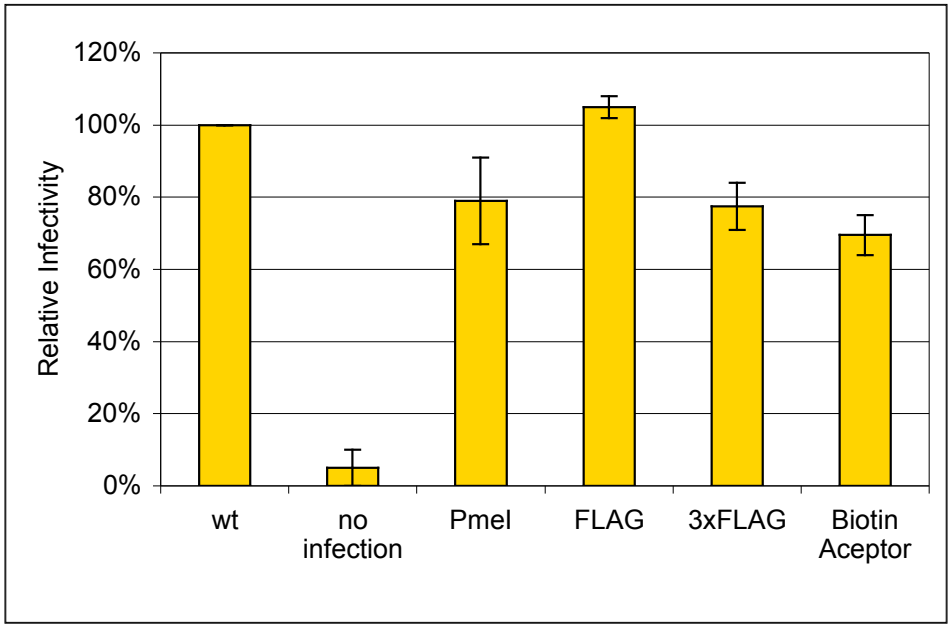


Figure 4.2 Viruses tagged in integrase are infectious in a single step assay. LuSIV (SIV:LTR-Luciferase) CEM cells were infected over a three log range of viral stock input normalized for identical amounts of p24 CA, incubated for 60 hours and then assayed for luciferase activity. The respective antigenic tag added to the integrase gene at the corresponding IN-PmeI site is indicated. The results are expressed as percent infectivity (wild-type = 100%) and are an average of two determinations within the linear range.

The integration frequency of viruses with PmeI, FLAG or 3xFLAG insertions within IN were also measured using a genetic assay based on the blastocidin D (*bsd*) gene carried in a single-cycle, VSV-G the pseudotyped virus. Integration frequencies of greater than 15 to 20% of that of the wild type are considered a requirement for sustained viral replication through several cycles in CEM T cell cultures (Wiskerchen and Muesing, 1995b). As shown in Figure 4.3, the R7/3-IN-PmeI, R7/3-IN-1xFLAG virus have an integration frequency of more than 50% that of the WT; the integration frequency of a virus with 3xFLAG tag was about 30% of the WT in this assay. Infections in other cell types such as GHOST(3)X4R5 and M7 5.25 were also studied. Results from these infections are consistent with the data described above.

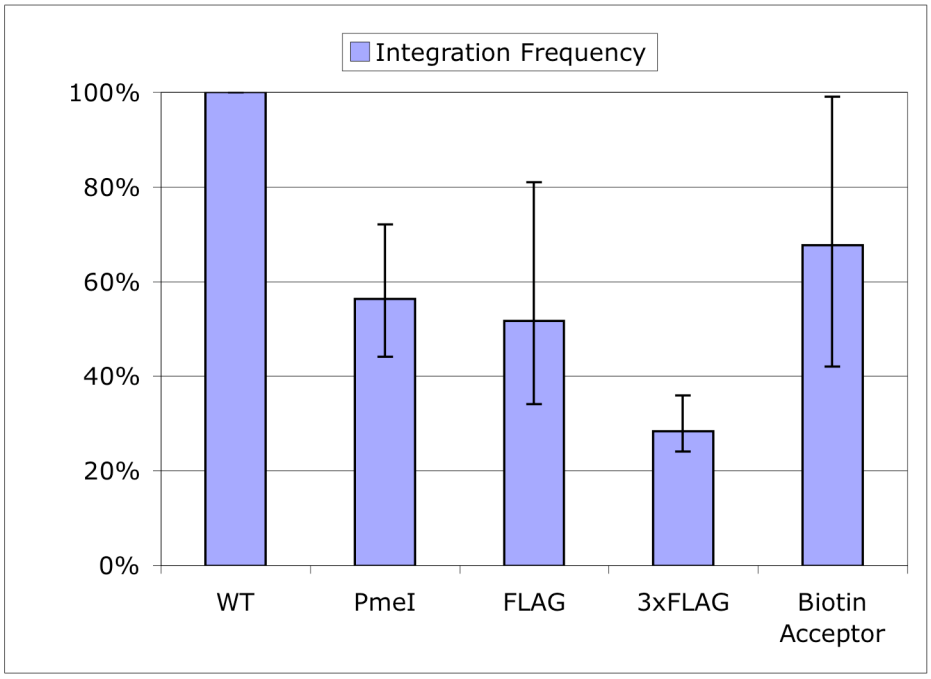


Figure 4.3 A genetic assay to evaluate the integration frequency of the various IN-tagged viruses. Data are an average of three readings.

4.2.2 Effect of V2 Insertion on Viral Neutralization Properties

We investigated the effect of foreign epitope insertion (i.e., the V2-3xFLAG modified Env) on the neutralization properties using a well-characterized monoclonal antibody active against most clade B strains of HIV-1. Since many of the monoclonal antibodies known to neutralize clade-specific HIV-1 species map to one of the variable loop regions, we asked whether insertional modification of the V2 region could still be presented as a neutralization target or whether this alteration disrupted the interaction of the antibody with its natural target. For example, insertion into the V2 loop region might alter the accessibility of a surface-exposed site such that the V2 loop or another loop-directed neutralizing antibody was no longer capable of binding gp120. To determine potential effects on neutralization, assays were performed with the known neutralizing mAb (447-52D) whose antigenic target is in the V3 loop (Conley et al., 1994). Neutralization assays of the wild type or 3xFLAG tagged Env were performed in 5.25.EGFP.Luc.M7 report cells, in which the infection could be monitored by the activity of either of GFP or firefly luciferase. As shown in Figure 4.4, mAb 447-52D effectively neutralized both the wild type and the V2-3xFLAG viruses. This indicates that at least at this approximation, the addition of 3xFLAG tag at the base of the V2 loop maintained the overall structure of gp120 and in particular, the structure of V3 loop itself. Not surprisingly, although the 3xFLAG-tagged Env virus is susceptible to anti-FLAG mAb neutralization, the wild type virus is not. Similar neutralization results were obtained when GHOST(3)X4R5 cells were used as the cellular target for neutralization and assayed by flow cytometry.

Thus, the insertion of the 3xFLAG sequence into the gp120 V2 region exerted no significant effects on the general neutralization sensitivity of the modified envelope glycoprotein, and by extension, indicates the possible neutralization of this specific recombinant virus by the activity of an α-FLAG mAb (M2).

However and interestingly, as shown in Figure 4.4B, the M2 antibody only partially neutralized this virus. In this experiment, a portion of the 3xFLAG tagged virus remained infectious despite incubation with high concentrations of the M2 antibody. Further increases in M2 concentration led to less specific virus inhibition, as evidenced by some effects of the FLAG antibody at higher concentration on viruses with the native wild type envelope glycoprotein (Figure 4.4B). It is therefore difficult to determine whether this residual fraction of infectious virus can be neutralized at the higher doses of M2 antibody without considering its inherent non-specific effects. Indeed, it is surprising that given the avidity of the M2 antibody for its target epitope, 3xFLAG, that the recombinant Env is not exquisitely sensitive to neutralization. It is possible that the M2 antibody does in fact bind to the 3xFLAG-V2 epitope in the context of the virion-associated Env trimer structure on the surface of the mature HIV-1 virion but the interaction does not lead to inactivation of the virus. On the other hand, the foreign epitope might be buried within the secondary, tertiary or quaternary structure of the mature envelope and inaccessible by antibody. Finally, since the modified 3xFLAG-tagged Env of this virus is quantitatively recovered from infected cells under native conditions, it is possible that the extensive post-

translational glycosylation of the viral envelope encountered during its progression from the endoplasmic reticulum to the surface of the infected cell and into the virion particle somehow occludes interaction with M2. It is perhaps significant that purification of this recombinant envelope from infected cells under standard immunoprecipitation conditions only recovers the gp160 preprocessed precursor protein (see below). Whether or not this protein is extensively glycosylated in the context of the foreign tag is currently unknown but serves as the basis for further investigation.

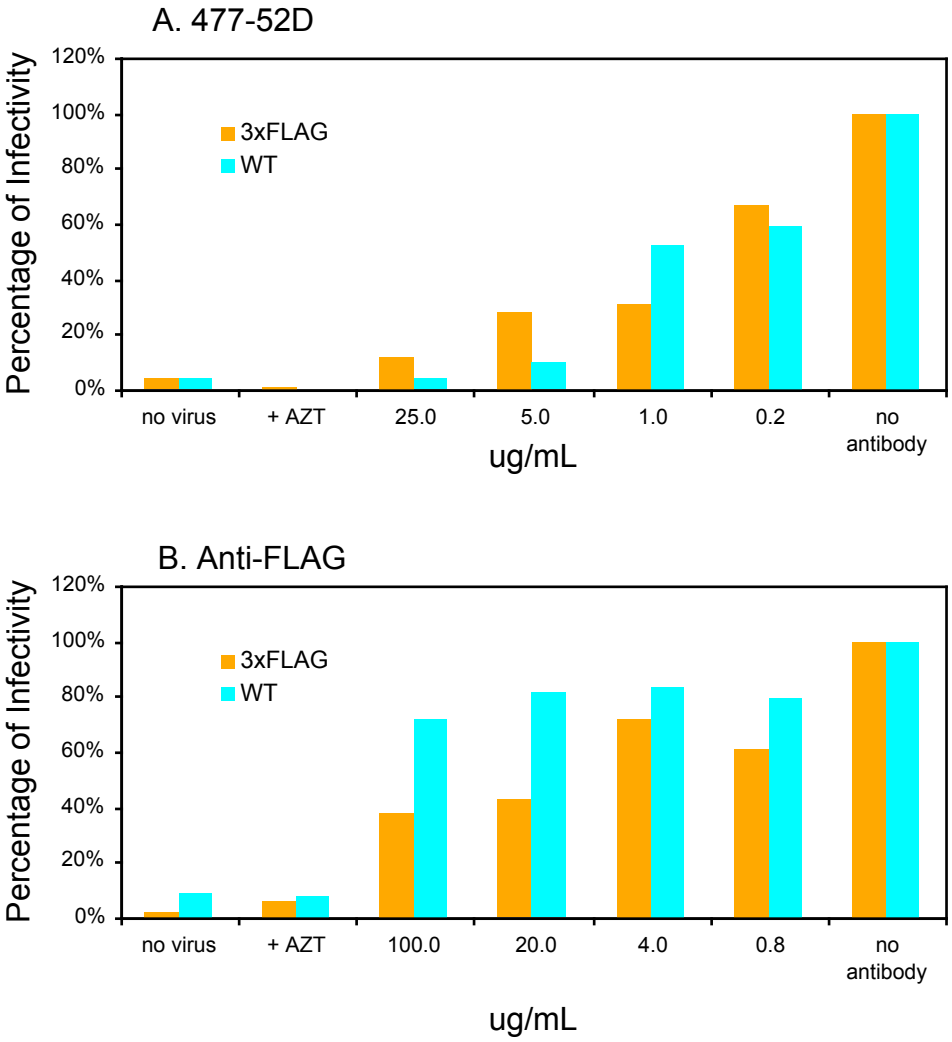


Figure 4.4 Neutralization assays using (**A**) 447-52D and (**B**) M2 α-FLAG. A mock infection with no viral addition and an infection with the RT inhibitor, AZT, serve as negative controls. A positive control was included in which neither antibody was added.

4.2.3 Visualization of 3xFLAG Tagged Viral Proteins During Infection

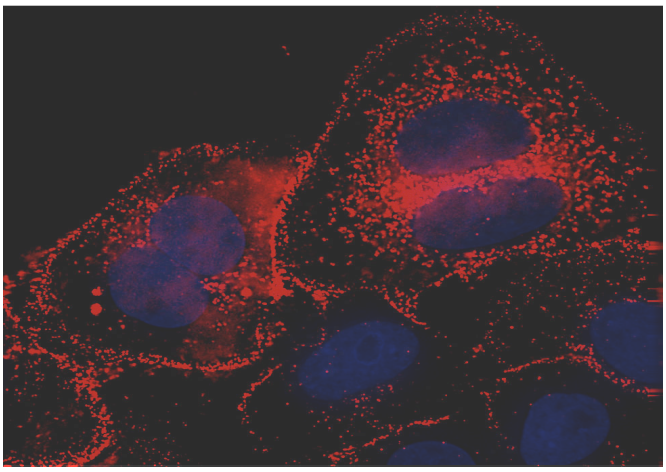
As an alternative but important utility, viruses with 3xFLAG insertions can be used to examine the subcellular localization of the tagged viral molecules during infection. HeLa P4R5 cells or GHOST(3)X4R5 cells were infected with different 3xFLAG tagged viruses and then probed with the M2 mAb and the complex visualized with the use of a fluorescein-conjugated secondary antibody. The cells were examined by fluorescence microscopy after 24-36 hours. As shown in Figure 4.5, all three 3xFLAG tagged viral proteins could be easily detected and visualized during infection (the heterogeneity in the signal intensity of individual stained cells being perhaps the reflection of positional effects on the specific location of the integrated proviral DNA). In cells infected with wild type virus, only very minimum background fluorescence was detected. For cells infected with the IN-3xFLAG virus, the majority of the signal was found in the cytoplasm as brightly fluorescing dots. The staining of the integrase-tagged protein (or the Gag-Pol polyprotein, of which integrase is a component) shows a reticular staining pattern surrounding the nucleus and extending throughout the cytoplasm. Its accumulation is also easily observed near the cell boundaries or plasma membrane (Figure 4.5 A).

Surprisingly, there was no obvious plasma membrane staining observed for virus with the 3xFLAG tag within envelope (Figure 4.5B). Instead, the tagged envelope protein was detected in the cytoplasm, concentrated near the nucleus. As discussed in the previous section, it is possible that only a subset of the 3xFLAG-

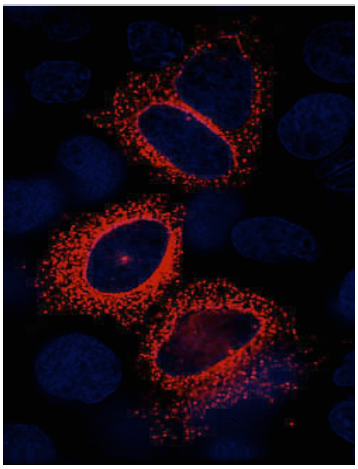
tagged Env protein is detectable by α-FLAG, perhaps reflecting either its state of post-translational modification or alternative structural configurations of the protein itself.

The staining of Vif-3xFLAG in the context of viral infection shows pan-cellular distribution (Figure 4.5C). The tagged Vif protein was detected in both the cytoplasm and nucleus without preference of accumulation in a particular subcellular compartment. The subcellular localization of Vif has been addressed previously (Goncalves et al., 1994; Karczewski and Strebel, 1996; Simon et al., 1997). These studies either investigated the expression patterns by transient transfection of a Vif expression vector or visualized at very low resolution during infection using a Vif monoclonal antibody during infection. Ectopic over expression of Vif was reported to be at the cellular membrane or within the cytoplasmic compartment. Our immunocytochemistry experiments visualize the punctate subcellular localization of Vif at high resolution in virus-producing cells. However, it is worth noting that since a Vif-permissive cell line was used in our study, the subcellular localization of Vif might vary based on infection of a Vif-permissive or restricted cell type. For our future study, we will investigate the Vif immunocytochemistry during infection of non-permissive cells.

A. R7/3-IN-3xFLAG



B. R7/3-Env-3xFLAG



C. R7/3-Vif-3xFLAG

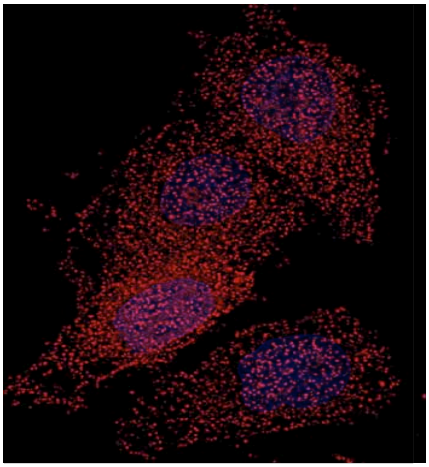


Figure 4.5 Immunocytochemical staining of HeLa P4R5 cells after infection with 3xFLAG tagged viruses. Cells were fixed and probed with anti-FLAG monoclonal antibody followed by Alexa-Fluor 594 anti-mouse as secondary antibody. The cells were counterstained for nuclear DNA (blue).

In an attempt to study the routing of the viral proteins in the viral life cycle, a time course with each of the tagged viruses was performed in infected GHOST(3)X4R5 cells. Since GFP expression in the GHOST(3)X4R5 cell line is conditional upon viral infection, these reporter cells were used to determine which of the cells in the population were infected. GFP can first be detected 18 hours after infection. As shown in Figure 4.6, we observed anti-FLAG staining for the IN, Env or Vif 3xFLAG tagged viruses at 18, 27 and 36 hours post infection. However, we did not notice any obvious redistribution of the tagged viral proteins. This may not be too surprising since by 18 hours most of the viral life cycle (approximately 24 hours) has been completed. At 36 hours post infection, syncytia formation, shown as large fused cells with multiple nuclei, was observed in all three infections. Further experiments with these viruses at shorter times after infection may be necessary for teasing out the intracellular movement of these proteins over time.

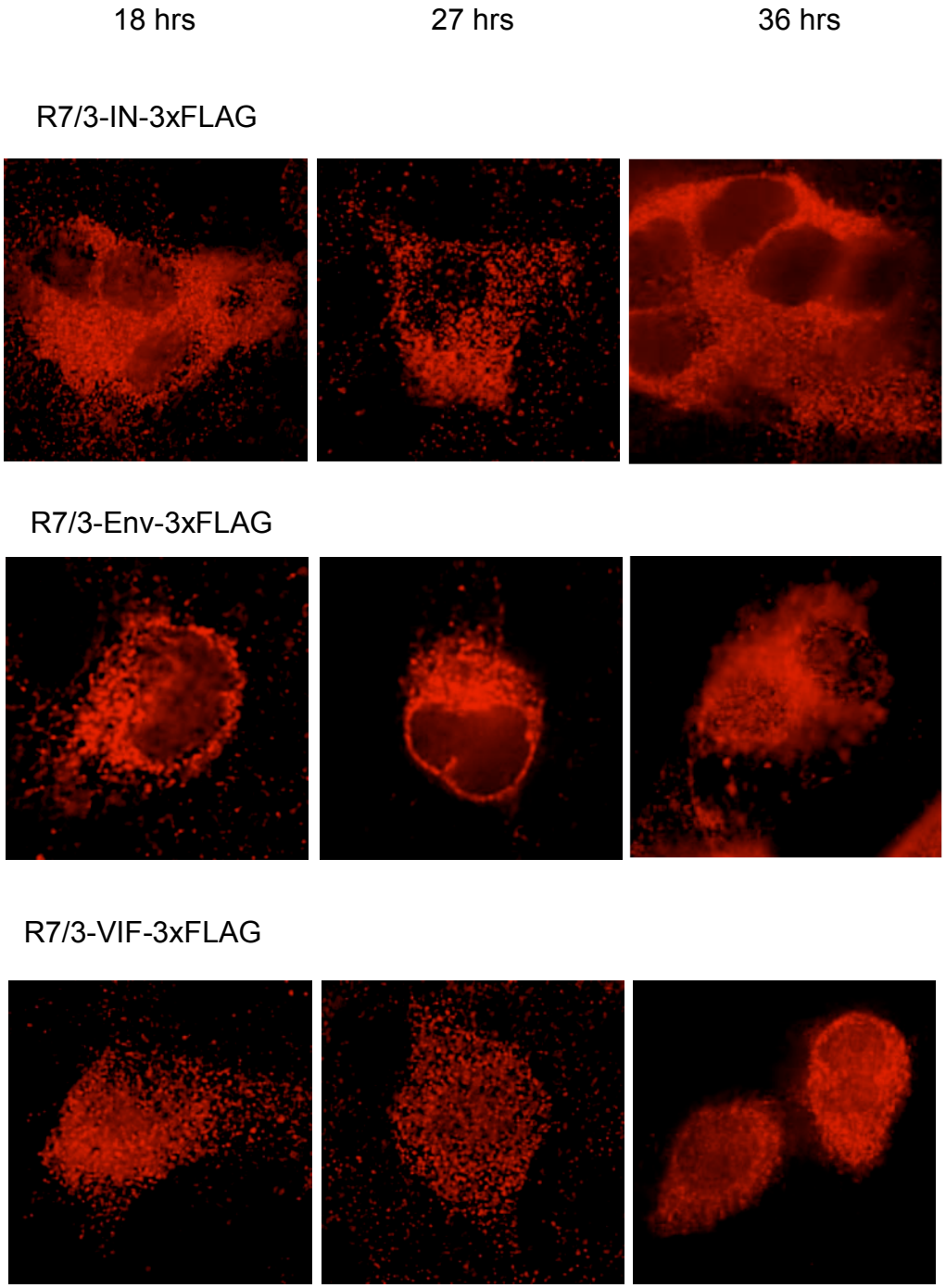


Figure 4.6 Immunocytochemical time course study of the location of 3xFLAG tagged proteins during infection in Ghost cells at 18, 27 and 36 hour time points.

4.2.4 3xFLAG Tagged Viral Proteins Can be Quantitatively Recovered During HIV-1 Infection

In pilot experiments, we assessed the ability of the 3xFLAG-tag to be an accessible substrate for the quantitative purification/recovery of the tagged viral protein during infection. Using the potent immunoaffinity characteristics afforded by the 3xFLAG-tag, whole cell lysates from infected cultures were incubated with α-FLAG(M2)-conjugated magnetic beads. To determine the extent of recovery, the eluted proteins from the beads were immunoblotted with the same FLAG antibody. We show that 3xFLAG tagged viral proteins were expressed efficiently in total cell lysates, (Figure 4.7A) and can easily be detected in their batch elution profiles (Figure 4.7B). Thus, all three viral proteins initially targeted for 3xFLAG-tagging were successfully recovered from progressive infectious cultures. All immunosignals were specific to each tagged viral protein and absent from the wild type control infection. Predictably, two major species of tagged proteins were recovered from the IN-3xFLAG infection—one the mature, fully-processed integrase protein (38 kDa), the other, the much larger proteins of molecular weights approximating 100 and 180 kDa. These proteins likely reflect the presence of the 3xFLAG tagged Gag-Pol polyprotein and related molecules, since integrase is part of this polyprotein before viral protease cleavage that occurs during viral egress from a producer cell. Under the conditions used in the pilot experiments, the recovered tagged protein in R7/3-Env-3xFLAG infections was at approximately 160 kDa, representing the gp160 precursor of the envelope glycoprotein instead of the mature 120 kDa gp120 glycoprotein species. Since

the mature form of gp120 is fully glycosylated and forms a homotrimer, the 3xFLAG epitope and α-FLAG binding might be less accessible in this form compared to the gp160 precursor (see discussion above).

Small-scale immunoprecipitation was performed using cells infected with single 1xFLAG tagged virus, R7/3-1xIN-FLAG, or the biotin acceptor tag, R7/3-1xSNAGG, the latter infected in cells expressing a biotin ligase. Importantly, although these viruses are more infectious than R7/3-3xIN-FLAG, to our surprise, the M2 α-FLAG mAb was unable to effectively isolate tagged IN or the Gag-Pol polyprotein from the infected culture even though the stable existence of the tag was confirmed by Western blot detection in the total cell lysate. Similar findings are observed for the biotin tag biochemical affinity purifications. It is quite possible that these smaller tags are buried in the folded viral integrase and not accessible for antibody binding.

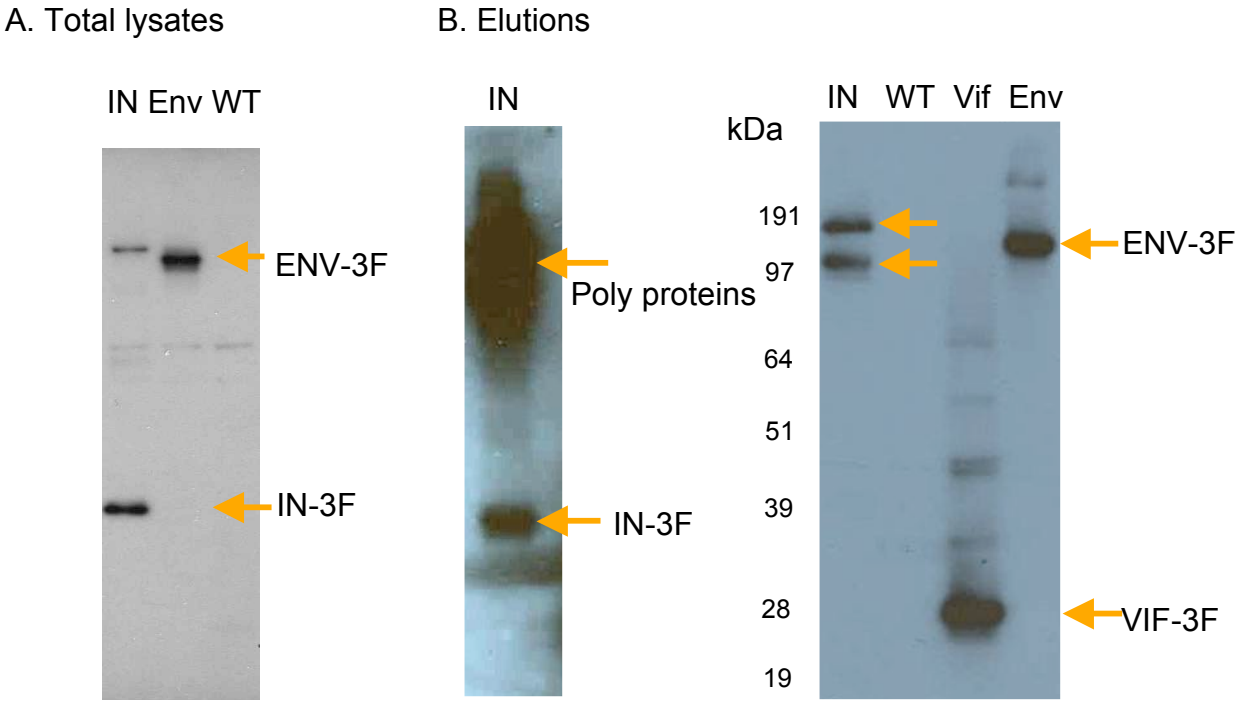


Figure 4.7 Immunoprecipitation of 3xFLAG tagged viral proteins during infection using α-FLAG. Western blots were probed with 1 μg/mL α-FLAG. Expression of the 3xFLAG in (**A**) the total lysate and (**B**) the eluted material is shown.

4.3 Discussion

The replication phenotypes of the viruses targeted for insertion were investigated quantitatively by growth curves and/or one-step growth analyses. Although the growth curves need to be reproduced in replicate experiments to include those of viruses with only the 5 amino acid PmeI site insertion as well as the replication defective viruses to be used as negative controls, these assays show that the growth kinetics of the viruses with 3xFLAG insertions is similar to wild-type R7/3, but with a slightly delay in viral production (Figure 4.1). As described in the previous chapter, the presence of the 3xFLAG-tag within IN, Env or Vif is stable and persists across many cycles of infection, modified viruses exhibiting wild type levels of viral protein expression and processing. Taken together, these results confirm the generation of replication competent viruses with long-term expression of a potent epitope tag. One-step infectivity assays for IN-tagged viruses (1xSNAGG, 1xFLAG, 3xFLAG) are consistent with regard to the results obtained over the course of a viral growth curve. In a genetic assay used to compare the integration frequency of viruses with various tag insertions, the 3xFLAG tagged virus shows an integration frequency in a range compatible with sustained viral replication in susceptible cells (Figure 4.3). Its lower integration frequency, perhaps indicative of slight perturbations in the efficiency of one or more functions carried out by integrase that are manifested when it is fused to a 30 amino acid exogenous and highly charged protein determinant. In general, it is not surprising that smaller tags result in less functional defects and are of less consequence to viral replication kinetics [illustrated by the behavior of the PmeI,

1xFLAG and the 1xSNAGG biotin acceptor viruses (Figure 4.2)]. However, although insertion of a larger tag into the same location might be expected to have more disruptive effects on the structure (or function) of the tagged protein and result in a reduction infectious viral titer, our results indicate that no insertion was dramatically more disruptive than any another at position 280 in integrase. This is not always the case. For example, the simple insertion of five amino acids (PmeI) at position Env 140 is replication-competent but when the 3xFLAG-tag is inserted at the identical position within this protein, the virus is severely compromised.

The relative replication fitness of various tagged viruses can be investigated by competition assay with wild type. This type of study could reveal which virus predominates when directly challenged in the same culture with the other viral species. In these experiments, cell would be infected with a 1:1 ratio of the tagged virus to the wild type virus. Proviral DNA can be extracted from infected cells and then amplified by PCR using primers flanking the relevant viral coding region. The ratio of the wild type to tagged virus, an indication of relative fitness, can be monitored daily by determining the ratio of intact to once cut (by PmeI or BlpI) in PCR fragment restriction digestion. An even more sensitive assay that can be used to estimate the relative fitness is based on growth competition experiments that use quantitative real-time PCR analysis (Weber et al., 2003). To quantify the proportion of both HIV-1 variants in the mixture, primers and probes can be designed specifically recognize and amplify the tag insertion or a specific

the viral coding region. The final ratio of the two viruses produced from each dual infection/growth competition experiment can be determined by real-time PCR and compared to production in monoinfections (Quinones-Mateu et al., 2000). A relative fitness value for each virus in the competition can then be estimated by the production of each individual HIV-1 strain in the dual infection.

Our data provide quantitative evidence for the overall rationale of our methodology – the biological selection of tags that exhibit relative inertness with respect to infectivity kinetics. The replication competency of the tagged viruses is a prerequisite of our strategy of host factor recovery and identification. Because of the amplification afforded by robust viral replication, the tagged viral protein represents a substantial proportion of the intracellular protein mass and thus is an enriched source for the discovery of binding events between viral and host proteins in the natural environments.

The HIV-1 gp120 variable regions are known to tolerate substantial changes in length, amino acid composition and glycosylation without compromising envelope function (Kuiken et al., 2003). We generated an infectious viral clone with a 3xFLAG tag at the base of the V2 loop region of gp120, an area thought to be unimportant for viral entry (Burton et al., 2004; Kwong et al., 1998). Thus, it is reasonable to ask if the 3xFLAG epitope is accessible for FLAG antibody binding and viral neutralization. In our experiments, the M2 monoclonal FLAG antibody could only partially neutralize the V2-3xFLAG virus. Although highly speculative,

the fact that the HIV-1 envelope can tolerate a large alteration in the V2 region indicates that this region likely plays a negligible role in the viral entry process. Since the overall structure of the envelope is maintained in this viral mutant, it remains susceptible to neutralization by conventional antibody targeting of the V3 loop. Therefore, the potential for neutralizing activity of the FLAG antibody is unlikely to result from disruption of a function intrinsic to gp120-receptor binding. Given the considerable size of an antibody molecule in comparison to that of the HIV-1 envelope trimer (Labrijn et al., 2003), steric hindrance may account for the reduced efficiency with which virus attachment and/or membrane fusion events occur in the presence of the bound antibody. Therefore, the neutralization activity of FLAG antibody might result from the binding and continued presence of the bulky antibody during the attempted entry process. It would also interesting to see if the antigen binding fragment (Fab) or the single-chain fragment V (scFv) of M2 FLAG antibody has the ability to neutralize the Env 3xFLAG tagged virus, given the fact that they are much smaller in size than the full-length antibody and may be more likely to avoid steric hindrance effects. There has been much debate on whether antibody binding to HIV-1 envelope glycoprotein necessarily leads to neutralization (Cavacini and Posner, 2004). Our study supports the notion that antibody binding with high affinity to Env exhibits some degree of neutralizing activity as a consequence of steric hindrance (Parren et al., 1998).

A fraction of the 3xFLAG tagged viruses could not be neutralized by α-FLAG mAb even at very high concentrations. At least two possible scenarios can be

envisioned: (1) The mature gp120 protein is highly glycosylated (Wei et al., 2003). Since the potency of neutralization by steric hindrance is expected to be largely determined by the accessibility of the epitope (Greenspan, 2001), the 3xFLAG epitope in a subset of viruses might be shielded by heavy glycosylation limiting antibody binding to its epitope. This hypothesis can be tested by introducing point mutations to remove glycan from Env and then studying the effect of this modification on neutralization. (2) Alternatively, although more unlikely than the scenario presented in (1), if binding between the 3xFLAG epitope and the antibody is reversible, a fraction of the virus may be able to bind to the cell membrane and elaborate fusion before the rebinding to the antibody. However, most of the evidence from our study seems to support the first possibility. The immunofluorescent staining of Env-3xFLAG infected cells only detects the cytoplasmic distribution of the mutant viral envelope instead of that found for its natural counterpart, where it is both cytoplasmic and accumulated at the plasma membrane, in position where the mature, fully glycosylated Env can be multimerized and incorporated into the virion. Furthermore, the majority of the envelope protein isolated during infection with the α-FLAG mAb is in the form of unprocessed gp160 (Figure 4.7). This indicates that the 3xFLAG insertion might be partially buried in the highly glycosylated mature gp120 or gp120 trimer, but accessible for FLAG antibody binding in the form of gp160, the unprocessed precursor of gp120. Although Env forms trimer on the surface of the viral particle, it is not clear if the epitope in each gp120 monomer need to be accessible for complete neutralization. Taken together, the reason for partial neutralization by

FLAG antibody is most likely due to the inaccessibility of 3xFLAG epitope in the mature oligomerized gp120 by its physical shielding either structurally or by post-translational glycosylation of the envelope protein.

Our study here shows that exogenous peptide tags can be introduced into the gp120 V2 region without affecting the recognition/fusiogenic function and conventional neutralization sensitivity of the HIV-1 glycoprotein. This is consistent with a previous report on the neutralization of a virus with 1xFLAG insertion in the V4 loop (Ren et al., 2005). Findings like this can be used for the establishment of an experimental system to study antibody-mediated neutralization, in which exogenous epitopes can be manipulated without interfering with normal structure and function of the carrier envelope glycoprotein. Such studies may provide further insights into mechanistics of HIV-1 neutralization by antibodies and other molecules.

Saturation mutagenesis of the HIV-1 envelope gene would possibly result in replication competent viral clones with 3xFLAG insertions at different locations. Neutralization of the viral infection by α-FLAG would be an indication that the particular 3xFLAG insertion is very likely in a surface exposed region of the gp120 trimer. Another application of the 3xFLAG tagged infectious clone in envelope is to investigate the effect of glycosylation on antibody neutralization. A proposed 'glycan shield' mechanism of neutralization escape whereby selected changes in glycan packing prevent neutralizing antibody binding has been

reported (Wei et al., 2003). It represents a new mechanism contributing to HIV-1 persistence in the face of an evolving antibody repertoire. By substitution mutagenesis on residues adjacent to the insertion, a virus with 3xFLAG tag insertion in Env that can be neutralized by α-FLAG can be used to establish that neutralizing antibody-selected alterations in glycosylation can confer escape from an epitope-specific monoclonal antibody.

We describe replication-competent HIV-1 that carry a potent 3xFLAG tag, which can, in practice, serve as a powerful tool to visualize any HIV-1 protein during infection. This strategy allows visualization of a functional viral protein species during viral infection in living cells, imaging providing a route to more fully understand the dynamics of viral host cell interactions. The HIV-1 viral protein R (Vpr) has been tagged with GFP to permit the first observation of intracellular HIV-1 complexes in living cells (McDonald et al., 2002). Recently, virus with a tetra-cysteine tag at the extreme C-terminus of IN was labeled with the bis-arsenical fluorescein derivative, FlAsH, for real-time imaging during infection (Arhel et al., 2006). Here, the location of the IN, Env or Vif proteins during infection was visualized by indirect immunocytochemistry. Since we have chosen to use the FLAG epitope for our recombinant viruses, we can rely on the commercial availability of a wide panel of anti-FLAG reagents (Sigma). For example, infections with our viruses can now be used for even more sensitive applications such as direct immunocytochemistry (fluorescent-conjugated) or electron microscopy (immunogold-conjugated) α-FLAG antibodies.

Our studies using indirect immunofluorescence reveal the subcellular localization of three tagged viral proteins during infection (or in some cases their immediate polyprotein precursor). We expected to see accumulation of the stained Env at cell membrane, where the viral budding occurs. As discussed earlier, the exclusive cytoplasmic staining pattern that we observed for the Env-tagged virus might represent the intracellular location of an early immature form of the gp160 precursor instead of the mature, fully-glycosylated form of gp120.

Although integrase is predominantly nuclear when expressed in the absence of other viral proteins, it elaborates a different staining pattern in cells producing the virus. It is both cytoplasmic and concentrated at the surrounding the plasma membrane (Figure 4.5A). Immunoprecipitation from a continuous infection with the IN-3xFLAG virus resulted in the overwhelming isolation of the Gag-Pol polyprotein. Since IN is present at the C-terminus of Gag-Pol, we suspect that the continuously infected cells from which immunoisolations were originally performed represent viral producer cells as a majority. As a result, the Gag-Pol polyprotein is preferentially recovered. On the other hand, the mature 38kDa IN proteins produced only after viral budding and in cells undergoing incoming infection , represent a minority of the total infected population in continuous, unsynchronized culture. However, when fresh uninfected target cells are added to the infected culture before harvesting (16-24 hours prior), there is considerably

more mature, 32 kDa IN immunoprecipitated than the Gag-Pol polyprotein precursor (see Chapter 5).

We show here the characterization of HIV-1 with tag inserted in *in*, *env* and *vif* gene products using various techniques such as infectivity and, neutralization assay and immunocytochemically. Our study is a starting point that may lead to versatile application of the modified viruses in an effort to understand HIV biology. Most importantly, the replication competent status and the accessibility of the tag in the infected culture provide a foundation for the use of our mutants to isolate viral protein and their cellular interactors directly from infection. This method is described in the next chapter.

Chapter 5: Identification of Cellular Proteins That Interact with Tagged Viral Proteins During Infection

5.1 Introduction

To efficiently precipitate HIV-1 viral proteins from infected cells and subsequently identify the cellular protein(s) they interact with, we isolated three infectious HIV-1 clones each carrying a 3xFLAG tag in IN, Env gp120 or Vif. These clones replicate efficiently in target cells when compared to the parental virus, R7/3. Because of the amplification afforded by viral replication, the tagged viral proteins represent a substantial proportion of the intracellular protein mass. Therefore, as cells infected with an individually tagged virus are expanded in culture they become an enriched reservoir for the recovery of viral-host proteins in complex. Cellular proteins that are associated with a particular viral protein under physiological conditions during HIV-1 replication can be recovered by immunoprecipitation via the strong avidity mediated by the 3xFLAG epitope tag. The purified protein complexes can then be identified by mass spectrometry (MS) and confirmed by a variety of techniques including tandem MS analysis, reverse targeting and I-DIRT (Tackett et al., 2005).

Strategies are needed to identify cellular proteins and their complexes that play a required role in HIV-1 biology. As discussed earlier, the multiple functions of HIV-

1 IN seemingly make this protein the most attractive target for our current study. However, this 'multi-tasking' ability strongly complicates the study of this protein. Several types of mutation in integrase are known have pleiotropic effects on disparate steps of virus replication and thus sometimes generate data that are difficult to interpret. For example, Q168A has been characterized as an IN single point mutation that blocks replication of the virus at the step of integration and thought to have direct effect on integrase catalytic activity. However, later study has shown that instead of affecting integration directly, IN Q168A actually disrupts the interaction of IN with the host factor LEDGF/p75, which tethers IN to chromatin prior to integration (Emiliani et al., 2005). In spite of recent efforts to elucidate the nuclear import pathway of the human immunodeficiency virus type 1 (HIV-1), the exact route(s) that the IN protein takes (as well as the domains that mediate its import) are still not understood (Armon-Omer et al., 2004; Bouyac-Bertoia et al., 2001; Yamashita and Emerman, 2005). Since IN is found in several different intracellular locations depending on the point of the viral life cycle, an additional problem in studying integrase may be that specific functions of the protein are determined by particular complexes formed with cellular proteins only in certain subcellular compartments (Van Maele et al., 2006).

Regardless of these complications, any study geared toward the identification of the components of host complexes associated with a particular viral protein will require their purification using a technique with high specificity and affinity. Here, 3xFLAG-tagged viruses were used to isolate host proteins complexes from

cultures infected with tagged, yet replication-competent mutants. We examined protein complexes involving tagged viral proteins produced during the infection of human T cells. In addition to some host proteins previously identified, mass spectrometry analyses of proteins recovered from tagged viral infections have revealed a larger number of interacting cellular proteins not previously described. These results provide the basis for a new way to understand of basic retroviral processes common to all retroviruses (i.e., proviral integration, envelope maturation) as well as those specific to the lentiviral family (i.e., nuclear localization) in which HIV-1 is the most studied member.

5.2 Results

5.2.1 Large Scale Immunoaffinity Isolation of 3xFLAG Tagged Viral Proteins From Infected Culture

Our strategy is illustrated in Figure 5.1. To follow interactions specific to the targeted HIV-1 protein, we used replication-competent viruses that could incorporate a 3xFLAG tag within the body of the IN, Env or Vif protein. As shown in Chapter 4, viruses expressing the 3xFLAG epitope are infectious with only slight delays in replication when compared to the wild type virus. As a negative control, immunoisolations were also performed with untagged wild type virally infected cultures to test for the presence of any nonspecific protein interactions that occur between the whole cell lysate and the specific antibody-bead

conjugate used in the recoveries. Infections by R7/3-IN-3xFLAG and R7/3-Env-3xFLAG in CEM target cells also served as ideal control for each other.

Infected CEM or Vif-restricted HUT78 T cells were rapidly frozen and cryogenically lysed to preserve protein complexes, as previously described (Cristea et al., 2006). The cell lysates were subsequently used for immunoaffinity purifications on magnetic beads coated with monoclonal α-FLAG M2 antibody, the purified proteins (and their complexes) identified by mass spectrometry after their purification by one- dimensional SDS-polyacrylamide electrophoresis.

Under optimized conditions, we have obtained efficient recoveries of the 3xFLAG tagged viral proteins from the infected cultures as shown by Coomassie Blue staining (Figure 5.2). The molecular weights correlate with the predicted sizes of the tagged proteins: 36 kDa for IN-3xFLAG, 163 kDa for the tagged, full-length Gag-Pol polyprotein, and 163 kDa for the glycosylated Env gp160 protein. The identities of the tagged proteins at these molecular weights were confirmed by MS analysis and Western blot. In this particular set of elutions, no band corresponding to Vif-3xFLAG (26 kDa) was detected by Coomassie blue staining. However, MS and Western blot analysis detected the presence of this protein.

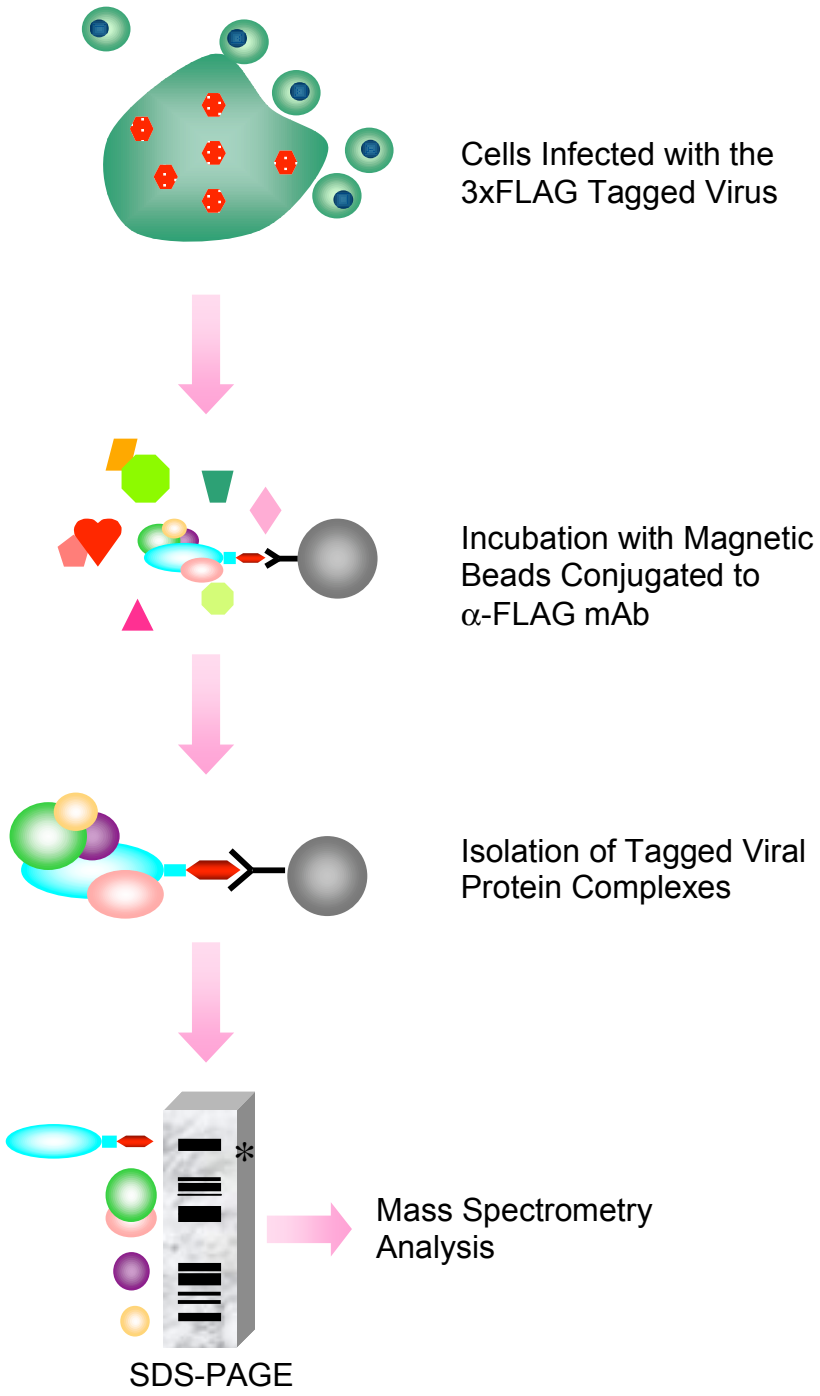


Figure 5.1 Scheme of the immunoaffinity isolation of tagged viral proteins and their complexes from infected cell culture.

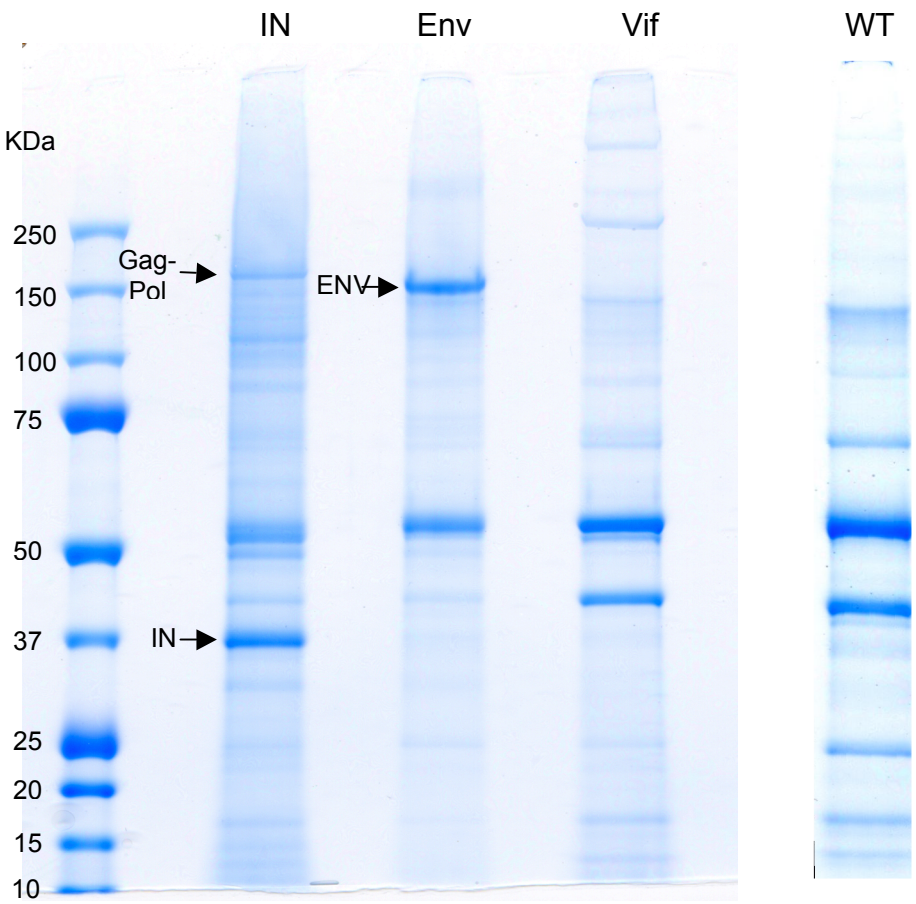


Figure 5.2 A preparative Coomassie blue-stained SDS-PAGE gel (4 to 20% Bis-Tris) of co-immunoprecipitated cellular proteins with either 3xFLAG tagged IN, Env or Vif. The bands confirmed as the tagged viral protein is indicated. The positions of molecular weight markers are indicated at the left.

The efficiency of immunoisolation was assessed by Western blot using the α-FLAG mAb. The eluted samples were compared to the fraction of the tagged protein in the total lysate and in the column flow-through. As shown in Figure 5.3, the expression of all 3xFLAG tagged viral proteins in infected culture were detected in the respective total cell lysate. In this experiment, the Env-3xFLAG protein appeared to be much more abundant than the IN-3xFLAG protein. However, based on detection of the tagged proteins in the elution lanes, most of the 3xFLAG tagged IN in the infected cultures was recovered and absent or diminished in the flow-through fractions. The additional bands in IN immunoisolation are indicative of the Gag-Pol or Pol polyproteins, or possibly a species of post-translationally modified integrase. Env protein was highly abundant in total lysate and saturated the magnetic beads for binding as indicated by the strong signal given in Western blot in the flow-through fractions (Figure 5.3). The tagged viral proteins were successfully eluted from the antibody-conjugated magnetic beads as evidenced by the absence of the tagged protein post-elution. In all of the eluted samples, the 3xFLAG-tagged proteins were present, their strong signal intensities on Western blot indicative of the high levels of tagged protein recovered after elution and correlating well with the results obtained from the preparative Coomassie blue staining of these samples.

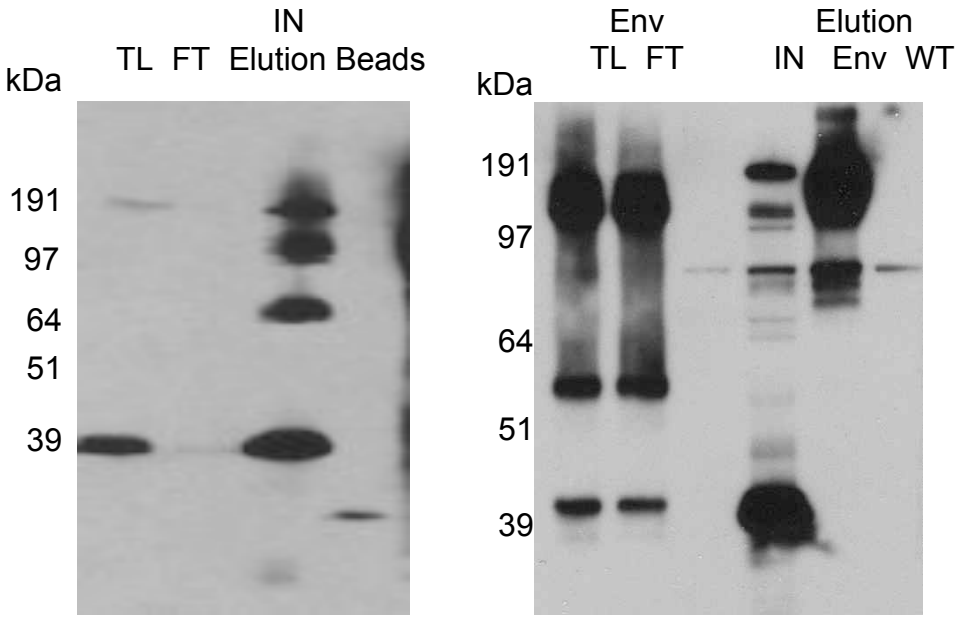


Figure 5.3 Immunoblot developed with α-FLAG to assess immunoprecipitation efficiencies. In the lanes for total lysates (TL) and flow-through (FT), approximately 0.1% of the sample volume was loaded; for elutions and beads recoveries, 1% of the sample was used.

5.2.2 Proteomic Characterizations of Proteins Recovered by Immunoaffinity Isolation During HIV-1 Infection

To identify the host proteins involved in HIV-1 infection, elution samples of each tagged HIV-1 immunoisolation was resolved by SDS-PAGE. Prior to electrophoresis, the cysteinyl residues within the samples were reduced and alkylated to yield carboxyamidemethyl derivatives and to avoid spontaneous intermolecular disulfide bond formation during the run. The eluted samples contained a large number of bands with a wide range of sizes (Figure 5.2). To analyze the proteins in the gel, each lane was sectioned into 24-26 contiguous slices. Each fraction was then digested individually with trypsin and the resulting peptides eluted from the gel by extraction. Peptide material extracted from each gel slice was first analyzed by MALDI-TOF MS, followed by MS/MS. The MALDI-TOF MS data confirmed that the digested samples contained peptides that could be assigned to cellular or viral protein. Searches of available databases indicate that the IN-3xFLAG and Env-3xFLAG samples contain statistically significant protein identifications. Thus, unique proteins presented in each of the immunoaffinity isolations appear to be from specific 3xFLAG tagged viral protein complexes and not from contaminating non-specific proteins. Some of the proteins identified in the IN-3xFLAG immunoprecipitation are divided into subsets of either cellular (Table 4) or viral proteins (Table 5). The genInfo identifier (gi) number, a unique integer that identifies a particular sequence, is included for each identified protein. Since IN and Env are most likely involved in different aspects of viral life cycle, cellular proteins that appeared in both IN and Env

samples probably represent nonspecific interactions. Some of them were eliminated as candidates for interaction since they are routinely recovered during immunoprecipitation and identified by mass spectrometry analysis from human cells (Cristea et al., 2006). The proteins that are commonly isolated through non-specific interactions are shown in Table 6.

Proteins recovered from the Env-3xFLAG infected cultures are displayed in Figure 5.5. Among the proteins recovered from the R7/3-Env-3xFLAG infection, the CD4 receptor was identified with a high level of confidence. Further confirmation of peptide sequences recovered from immunoprecipitation samples are in the process of being analyzed by tandem MS analysis. So far, we have confirmed MS identification by comparison of parallel mobility fractions in which wild type infected but untagged control cultures were contrasted with IN or Env protein recoveries.

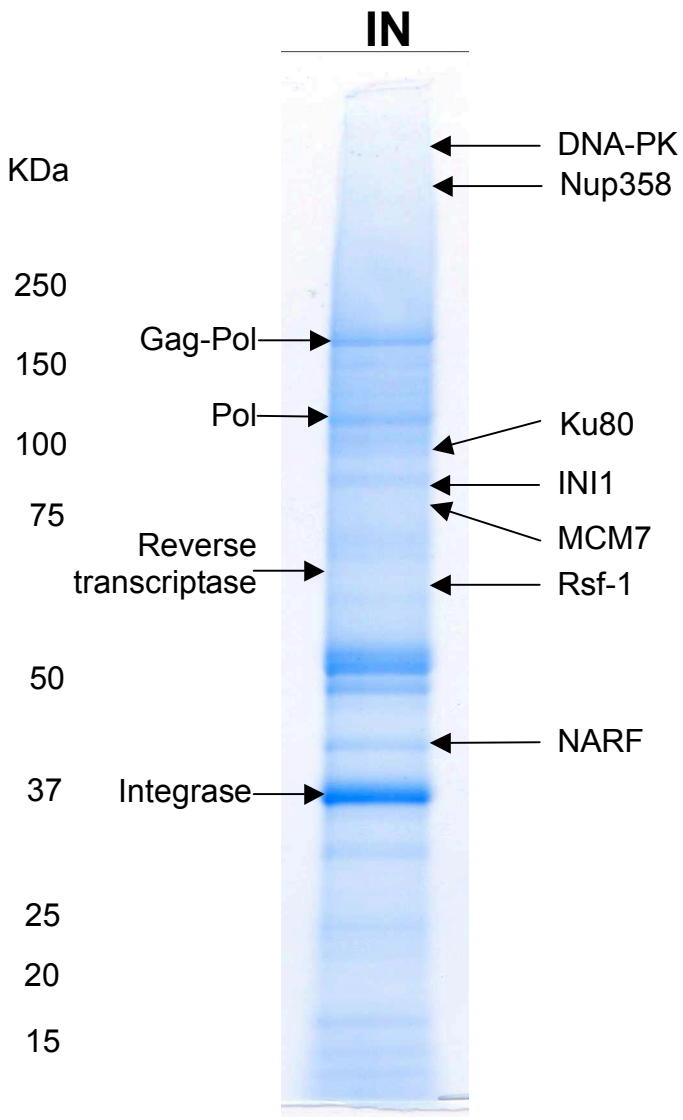


Figure 5.4 The mobilities of proteins recovered from the IN-tagged immunoprecipitation. In this figure, viral proteins are shown on the left of the gel lane, the recovered cellular proteins are on the right.

Table 4. Cellular proteins identified from the IN-3xFLAG immunoprecipitation.

Function Involved	**Protein Name**	**GI Number**
Nuclear Import	Nup358	gi|83305554
	RanGAP1	gi|4506411
	TPR	gi|114155142
	POM210	gi|97180292
Ubiquitination	HECT E3 Ub ligase	gi|51094760
	AN1	gi|28376664
	NARF	gi|12653339
DNA Repair	DNA-PK	gi|113430845
	ATP-dependant DNA Helicase II (Ku80)	gi|10863945
Chromatin Remodeling	SWI2/SNF2	gi|15529034
	MN1	gi|94730396
	SNF5 (INI1)	gi|27545326
	Hsp70	gi|62896815
	MCM7	gi|755746
	Rsf-1	gi|388788333

Table 5. Viral proteins immunoprecipitated with 3xFLAG-tagged IN during viral infection**.** The expected molecular masses were within the range of their observed masses.

Protein Name	**gi number**	**Molecular Weight (kDa)**
Gag-Pol	gi|94732061	160
Pol	gi|38491490	90
Gag	gi|82568630	55
RT	gi|49574314	66, 51
Nef	gi|47779058	27
Env	gi|46882803	120 (size of gp120)
Vif	gi|32399659	23
p24 (CA)	gi|3183828	24
Protease	gi|24021494	22

Table 6. Proteins (Homo sapiens) that are obtained in all lanes and considered as common contaminants in all three immunoisolations.

Protein name	**gi number**
Myosin VC	gi|9055284
Tubulin beta polypeptide	gi|5729813
Vimentin	gi|340234
Dynein heavy chain domain 3	gi|75677365
Beta actin	gi|16924319
Plectin	gi|41322910
Keratin	gi|7717238

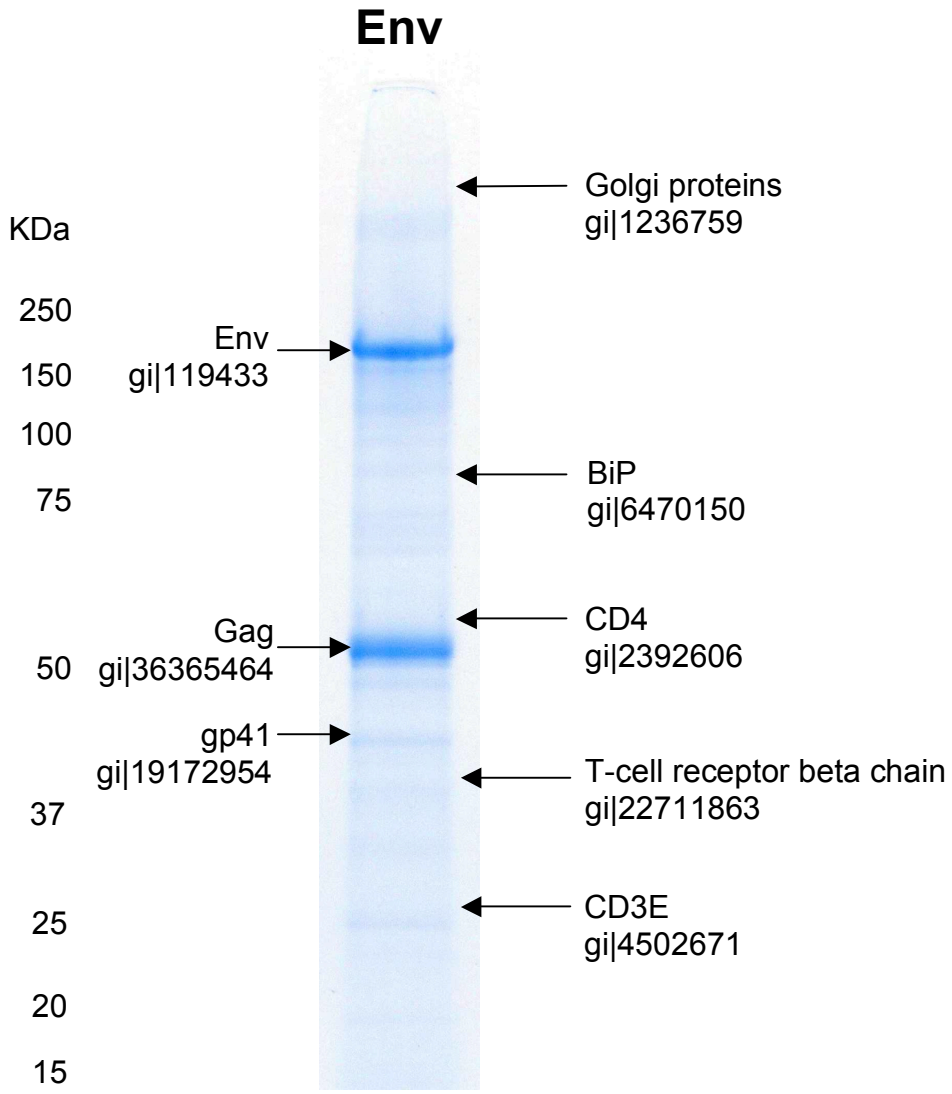


Figure 5.5 The identity and mobility of the proteins recovered from the Env 3xFLAG tagged immunoprecipitation. In this figure viral proteins are shown on the left of the gel lane, the recovered cellular proteins are on the right,

5.2.3 Immunoblot Analysis of Virus-host Interactions

A series of pilot experiments were initially carried out to test or confirm whether specific cellular proteins could be co-immunoprecipitated with 3xFLAG-tagged viral proteins during infection. Immunopurified samples were examined by Western blot analysis using specific antibodies against the cellular proteins. Previous small scale immunoaffinity purification experiments of virus with the IN-3xFLAG tag did not detect the presence of some cellular proteins documented with various certainties to interact specifically with integrase such as LEDGF (Cherepanov et al., 2003), p300 (Cereseto et al., 2005) or Gemin2 (Hamamoto et al., 2006) by immunoblotting.. However, as shown in Figure 5.6, specific interaction between the envelope protein and its CD4 receptor was revealed in a small-scale immunoprecipitation, in which 10 mL continuously infected culture was used under a lysis condition of 50 mM Tris-HCl, 150 mM NaCl and 1% Triton. Elution from wild type infected culture was used as control. It should be noted that these pilot experiments did not entail the cryogenic lysis procedure as outlined. In addition buffer system was used in these recoveries.

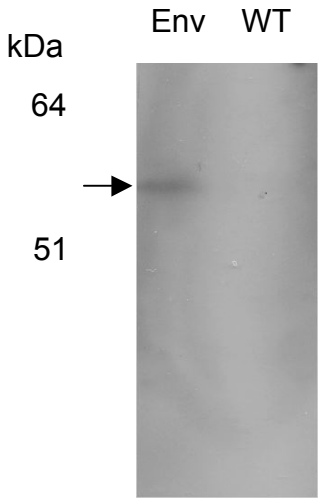


Figure 5.6 Immunoblotting of the Env-tagged HIV-1 immunoisolation using the wild type infection in parallel as the negative control. Sample identity is indicated. This blot is probed with the α-CD4 mAb, OKT-4.

5.3 Discussion

The study of cellular proteins involved in the HIV-1 life cycle has been complicated by the lack of antibodies with high binding affinities and specificities against the proteins of HIV-1. Here, we describe the development of a robust strategy that uses modified infectious virus harboring a potent antigenic tag within a target viral protein to significantly enrich the recovery of the tagged protein along with its associated host protein complexes. Importantly, most contaminants from the infected culture lysates are removed quickly in a single step that precedes their rapid identification. The immunoisolation method we use is based on the studies of our collaborators that use GFP as their proteomic probe (Cristea et al., 2006; Cristea et al., 2005). The rapid freezing and cryogenic grinding procedures outlined in these papers uniquely preserve protein complexes close to their original state in the cell. The buffer conditions used here have been previously optimized for the immunoprecipitation study of Sindbis infection. We show here that this proteomic recovery system can be adapted to more size-restricted genomes like HIV-1 if a high affinity epitope tag is used to facilitate rapid single-step immunoaffinity isolations. In turn, this procedure minimizes nonspecific binding while maximizing the recovery of even transiently interacting cellular partners.

To capture the viral-cellular complexes encountered during infection, the 3xFLAG tagged IN, Env, Vif, as well as wild type infected target cells were prepared from large, one-liter cultures of susceptible cells. The yield of the Vif immunoisolation

was low probably due to the targeted cell type used. The Vif-restricted cell type HUT78, which does not form syncytia during infection, was used for the immunoisolation experiment. Since the infection could not be monitored by following syncytia formation, it was possible that only a small portion of the cells were infected. The IN-3xFLAG and Env-3xFLAG can serve as controls for each other since they are completely different proteins but both tagged with the 3xFLAG peptide. In fact, even though the tagged viral proteins normally reside in different subcellular compartments during infection, the cryogenic methodology used here is capable of complete mixing of the freeze-fractured compartments prior to thawing in lysis buffer. Thus, the opportunity for possible non-specific contamination of the α-FLAG-conjugated beads should be nearly identical regardless of the tagged viral protein. This makes comparison between proteins isolated from the IN or Env tagged viral protein purifications discriminatory with respect to their specific interaction—nonspecific interactions considered to be present in both elution samples. Many highly abundant cytoskeletal proteins were observed in both immunoisolations, such as microtubulin and actin. They are often encountered during mammalian lysate isolations as nonspecific binding proteins (Cristea et al., 2006), although we cannot rule out the possibility that these proteins are involved in the directional intracellular transport of viral contents (Fackler and Krausslich, 2006). On the other hand, a study using an anti-idiotypic antibody to the V3 domain of gp120 has indicated that vimentin may participate in the early steps of HIV-1 replication (Thomas et al., 1996). We noted

that several proteins like vimentin and myosin are present in both samples raising questions as to the specificity of the reported interactions.

The proteomic analysis of HIV-1 immunoprecipitation from the tagged viral infections reveals many candidates that have not previously been described to interact with these viral proteins. The host proteins identified in our preliminary experiments are components of a wide variety of cellular systems encompassing a wide spectrum of intracellular pathways. Since the limited genomic complexity of viruses in particular dictate their substantial reliance on a vast array of different cellular systems for their survival, the identified cellular proteins, after verification for specific interaction, should provide potential clues for their role in viral biology. Table 4 provides a list of the proteins identified to interact specifically with tagged integrase during infection. Of the unique proteins identified here, some have been previously been reported to associate with HIV-1 integrase. SNF5, also known as integrase interactor 1 (INI1), was the first cellular protein shown to directly interact with integrase (Kalpana et al., 1994). The ATP-dependent DNA helicase II, also known as Ku80, has been reported to associate with the viral preintegration complexes containing HIV-1 IN (Li et al., 2001). DNA-dependent protein kinase (DNA-PK) was also identified in our search. The requirement of the DNA-PK for IN-interaction activities is controversial. It has been proposed that, DNA-PK, along with Ku70 and Ku80, play a role in retroviral DNA integration and protect cells against toxicity induced by HIV-1 Integrase or the integration

process itself (Daniel et al., 1999). However, other studies have suggested that DNA-PK is not required for efficient lentiviral integration (Baekelandt et al., 2000).

The primary integral plasma membrane protein required for HIV-1 entry (CD4 receptor) was among the cellular proteins found to interact with Env-3xFLAG during infection. This indicates that Env 3xFLAG maintains its ability to bind to this cellular receptor and is consistent with a requirement of this cellular factor for the growth and expansion of the Env-tagged virus in culture.

The identification of known interactions supports the rationale of our approach. Traditional methods for the isolation of protein complexes are usually based on expressing the protein of interest as tagged by N- or C-terminal fusion in the absence of other viral proteins. This strategy, however, does not always represent the natural situation for interaction. For HIV-1, the more biologically relevant way to express a viral protein is to produce it in the context of viral infection. However, in this case, the location of tag insertion becomes critical. Although 1xFLAG tag insertion at the C-terminal end of IN and insertion of 1xHA tag at the C-terminus of Vif has resulted in replication-competent virus (Petit et al., 1999; Yu et al., 2003), one cannot expect this outcome for all viral proteins. The cloning of a tag at the N- or C-terminus might affect polyprotein processing. In addition, over the past 8 years since the initial publication of the 1xFLAG-IN virus, no reports using this virus for proteomic studies have appeared in the literature. The reason for lack of its use may be reflected in data from our

laboratory using a related 1xFLAG tagged virus at the identical position (IN 280) as in our valuable 3xFLAG-IN virus. Using this 1xFLAG-tagged virus, we have been unable to quantitatively recover IN from infected cells (see above).

We developed the transposon-based insertion strategy to select at random, those sites in the targeted viral protein that can tolerate an insertion. This procedure is without human bias and allows the virus itself to self-select the most inert position for insertion. We feel that the strategy employed here is an efficient and potentially powerful approach that is generally applicable to the study of other viral proteins.

As shown in table 4, proteins eluted from 3xFLAG-IN viral infections contain many proteins from several important cellular systems, including the nuclear pore complex, the ubiquitin/proteasome system, the DNA repair and the chromatin remodeling assemblages. In addition, we have also identified membrane-associated proteins, an RNA export factor and ADP/ATP carrier. However, at present, the relevance of these proteins to viral biology is unclear. While most of these proteins provide important possibilities for influencing viral function, the present observations only represent a starting point for further extensive analyses. MS/MS and I-DIRT will be used to confirm the peptide sequences and verify the presence of the cellular proteins in the immunoisolations. Further studies such as co-localization, reciprocal immunoprecipitation and RNAi will be needed to determine the relevance of these host proteins to viral life cycle.

The I-DIRT technique can be used to determine whether a given interaction is specific (Tackett et al., 2005). Briefly, cells infected with the replication-competent virus expressing an affinity-tagged viral protein are grown in light isotopic medium while wild-type cells or cells infected with wild type HIV-1 are grown in heavy isotopic medium such as DL-lysine-4,4,5,5-d4 2HCl (d4-lysine). (The molecular weight difference between the two isotopes is 4 Daltons.) Equal amounts of the two preparations are then mixed and the isotopically light affinity-tagged protein is immunoisolated. Both specific and nonspecific interacting proteins will be recovered with the affinity-tagged protein. Interactions that are specific will form with the tagged protein prior to cell lysis, and will be exclusively light labeled. In contrast, nonspecific proteins, which form their association with the tagged protein after thawing of the lysed mixture of light and heavy cells, will be both isotopically heavy and light labeled. After isolation of the affinity-tagged protein complex and trypsin digestion of the co-enriching proteins, specifically interacting proteins are identified by mass spectrometry as isotopically light, while nonspecifically interacting proteins appear as a mixture of light and heavy isotopes.

We believe that abundant and high affinity binding cellular proteins are more likely to be immunoisolated with the tagged viral proteins from infection. The buffer conditions used for our large-scale immunoprecipitation (20 mM K-HEPES, pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 0.1% Tween 20, 1% Triton, 0.5%

deoxycholate, 0.3 M NaCl, and 25 U/mL DNase) were originally optimized for Sindbis immunoprecipitation (Cristea et al., 2006). Buffer conditions are critical having effects on the subset of cellular proteins co-immunoprecipitated with the tagged viral proteins. For example, binding of IN to LEDGF has been assayed *in vitro* in 25 mM Tris-HCl, 0.1% NP-40, 20 mM imidazole containing 150 to 400 mM NaCl in the presence of 1 mM $MgCl_2$ (Busschots et al., 2005), a much milder set of conditions than the ones used in our study here. A panel of detergents as well as different salt concentrations can be used to optimize the 3xFLAG tagged viral protein immunoprecipitation from the HIV-1 infected cells. It is possible that under certain modified buffer conditions, additional interacting cellular factors including other known integrase binding partners such as LEDGF and Rad18 may be revealed.

Meanwhile, infection of different cell types might result in an overlapping but distinct set of cellular proteins recovered during immunoisolation. For example, infection of primary macrophage cells or other non-dividing cell types may reveal factors involved in active nuclear import of the virus, perhaps being more abundant in these cells.

In our study, only one 3xFLAG clone has been selected for each of the original target proteins. It is quite possible that viruses with tags positioned at different locations within a viral protein would also be infectious. Recovery of the same set of host factors from tags placed at different locations within a target protein would

provide compelling evidence for authenticity of the interactions. Conversely, while most of the set of recovered host proteins would be expected to be the same, it is also possible that a few cellular proteins might be recovered with only one of the tagged proteins. Genomic fingerprinting can assist in the identification of new sites in the targeted protein most amenable for further insertional modification.

There are several caveats to be considered when interpreting the data. First, the tagged viral proteins were recovered from continuously infected culture, which is a mixture of all stages of the viral life cycle after seven days of continuous propagation. Therefore, particular host protein recovered might only represent a negligible portion of the whole infected culture. Second, the expression level and variety of cellular proteins can change during the course of infection. Interaction observed at a certain point of the viral life cycle might result from differences in the expression levels of cellular proteins. Another consideration is that while the MS analysis used here identified many proteins, some known cellular partners were not detected. For example, the lens epithelium-derived growth factor (LEDGF/p75), a factor known to interact with HIV-1 integrase (Cherepanov et al., 2003; Maertens et al., 2003), was not detected in the IN elution. Failure to identify a particular protein by MS does not necessarily indicate that the protein is not present in the analyzed sample, since the absence of identifiable peptides can be due to several technical factors. In this case, because the samples were treated with integrase inhibitor prior to lysis, some protein interactions might be partially blocked in the presence of this compound. This reflects the intrinsic

difficulty in identifying proteins present in relatively small amounts within complex mixtures, even after large-scale purification. Alternatively, since the IN-LEDGF interaction was detected only upon physical cross-linking and in the context of an IN domain yeast two-hybrid search, the LEDGF-IN interaction may not be of sufficient strength to be detected here.

There are a few limitations of this strategy if it is applied to other HIV-1 genes. HIV-1 has several overlapping reading frames, for example, *pol* overlaps with *vif*, *env* overlaps with *rev* and *tat*. Insertions into the overlapping region of one protein would necessarily result in the mutagenesis of the other. In the case of integrase, this problem was alleviated by using Vif-permissive target cells. However, to circumvent this problem in general, cells that constitutively express one of the viral proteins can be used as the target cells for infection or expressed in cells assembling the virus in either case to preserve the function of the particular viral protein.

Our MS analysis has identified cellular proteins potentially involved in various aspects of virus replication, but also identified proteins whose presence is difficult to interpret based on current knowledge. Several explanations are possible. First, the cellular proteins have unknown functions, or multiple pertinent functions other than which they are best known (Jeffery, 2003). Second, some of the proteins identified here also could be simply adhered to the viral protein nonspecifically and are not functionally relevant. However, this proteomic analysis represents

important data because it provides the basis for further studies aimed at the involvement of cellular proteins and their pathways in the HIV-1 replication cycle. Such studies will provide additional information about the various pathways used by HIV-1 in infected cells, raise interesting questions about the possible contributions of these viral assisting molecules in HIV-1 pathogenesis, and may eventually point to potential cellular targets for novel approaches to therapeutic intervention of HIV-1 infection.

The electrophoretic mobility of many of the observed proteins from the preparative gel corresponds to their expected size as deduced from their amino acid sequences. For example, Ku80 was found in the gel fractions corresponding to 80 kDa. However, some of the proteins were identified outside of the normal molecular weight range. For example, IN was found in range 20 kDa to 50 kDa. However, this might represent a special case because IN is also a component of the large (183 kDa) Gag-Pol protein and other of its intermediate proteolytic intermediates that are larger than the mature IN molecule. Post-translational modifications, such as phosphorylation, ubiquitination or glycosylation may cause proteins to migrate more slowly than expected and to be found in higher molecular mass bands on the SDS gels. On the other hand, some proteins migrated at a faster rate, reflecting a lower molecular mass than expected for the full-length proteins. The presence of truncated forms of cellular proteins might be due to their degradation, alternative splicing of their mRNA or post-translational cleavage by non-specific cellular proteases.

MS also detected several viral proteins from the IN-3xFLAG immunoprecipitation. This list includes Gag-Pol, matrix (MA), capsid (CA), p51 and p66 reverse transcriptase (RT), protease (PR), Nef, Vif and p24 (Table 5). As described in the previous paragraph, the biological significance is potentially masked by the fact that during its maturation, IN is expressed as in integral portion of the Gag-Pol polyprotein, only cleaved by HIV-1 PR after egression of the virion from the producer cell (Vogt, 1996). On the other hand, Gag processed products (MA and CA), RT and IN are thought to be part of the RTC/PIC. Therefore, it is not possible to distinguish between these two possibilities in the current experimental setting. Immunoprecipitation experiments at different time points synchronized after infection can be used in the future to address this question.

During the development of our strategy, various tags were tested including Strep, 1x- and 3x-FLAG and a biotin receptor tag (SNAGG) where biotin can be added co-translationally by co-expression of the bacterial *bir-a* gene in cells. The 1xFLAG tag is considerably smaller than the 3xFLAG tag and its addition results in viruses that are less compromised for replication. However, the binding affinity of the M2 α-FLAG mAb for 1xFLAG is at least 1-2 orders of magnitude weaker than it is for 3xFLAG. The strep-tag and the SNAGG-tag utilize the extremely strong binding between biotin and streptavidin (Wilchek and Bayer, 1990). They have the advantage of being relatively uncharged, neutral and smaller compared to the acidic 3xFLAG tag, and therefore, perhaps more likely to be less

structurally disruptive. Surprisingly, these tags were both less efficient and less specific during immunoaffinity isolation. It is possible that the shorter tag is hidden, buried or unavailable for antibody or binding partner accession. It has also been documented that some engineered biotinylated molecules often bind to streptavidin with affinities that are reduced by several orders of magnitude, compared to the extremely high affinity of native biotin for streptavidin (Chilkoti et al., 1995).

The abundance and the affinity of the strength of the binding are both critical parameters for successful immunoprecipitation results. Infections of cell types other than T cell lines were examined for tagged protein recovery. For example, GHOST(3)X4R5 and HeLa P4R5 cells were used because these adherent cell types can be efficiently infected and used to detect the intracellular trafficking of the tagged viral protein after infection. We also tried a cross-linking treatment were performed with the tagged viruses to probe weak or temporary interactions. The VSV-G envelope was used to pseudotype tagged virus as it is more infectious than either the X4- or R5-tropic HIV-1 Env and allows entry into almost all mammalian cell types at high m.o.i. However, the VSV-G pseudotype strategy can only be used for a single round of viral infection. In the end, the best system for proteomic studies is the infection of T cell lines with a 3xFLAG tagged virus specifying the natural X4-tropic HIV-1 Env protein. This approach clearly results in the most available substrate for immunoprecipitation by virtue of the amplification and expansion of the virus over a several day period. Thus, the conditions to obtain the most prodigious viral expansion during T cell infection

'scale-up' have been optimized to obtain the largest amount of the targeted viral protein for our studies.

The keys to the success of our approach include the random introduction of a potent, yet relatively small, tag within a specific HIV-1 viral coding region, the incorporation of a stringent selection for the viruses that can accommodate the tag yet remain infectious in culture and the use of state-of-the-art tools to identify host proteins associated with viral proteins after their recovery from infected cells. Our study of proteins that interact with HIV-1 IN, Env and Vif during natural infection represents a "breakthrough" technology that can be further developed to identify cellular factors that associate with other viral proteins. The work presented here only represents a first step in the comprehensive description and characterization of the intricate interactome that must exist at the interface of virus and host. As an extension of this approach, animal models can be studied using a similar approach to investigate virus replication, spread, and pathogenesis *in vivo*. Of course this represents a higher degree of difficulty since success in an animal system not only depends on the ability to insert a tag within a protein without disrupting its function but most importantly, long-term stability of the recombinant tag.

A similar strategy has been reported to study Sindbis virus *in vitro* using replication-competent GFP-tagged viruses (Atasheva et al., 2007; Cristea et al., 2006; Navaratnarajah and Kuhn, 2007). We have considered use GFP protein as

a possible insertion since the GFP tagged viral protein can be visualized by the green fluorescence during infection. However, the 27 kDa GFP is a relatively large insertion compared to the compact genome of HIV-1. An attempt to insert GFP into the PmeI site within the HIV-1 V1-V3 region of envelope was not successful. Independently developed, our system is the first one for the study of HIV-1 proteins and overcomes the major hurdle of intergenic recombination. Therefore, we predict that the presently described technique will be broadly applicable to many viral systems and will greatly facilitate our understanding of the molecular details of the complex interplay between the virus and the host during pathogen establishment and persistence.

Implications of Current Study and Prospect for the Future

We have successfully developed a system to molecularly engineer HIV-1 in such a way that the incorporation of a foreign immunological tag does not appreciably affect sustained viral replication in culture. Using independently tagged replication-competent derivatives we have recovered host proteins that interact specifically with the tagged viral protein from continuous infection. Mass spectrometry has generated a list of potential interacting partners for each viral protein. In addition to reproduce the large-scale immunoprecipitation from infected material and the mass spectrometry analysis, several methods will be used to validate the identified host factors. Immunocytochemistry of the interaction will be studied. A coherent pattern of subcellular localization between the viral and host proteins may suggest the association of the proteins in the intracellular environment. For immunofluorescence studies, cells will be co-transfected with an epitope-tagged cellular protein and its putative viral ligand. Alternatively, the host protein subcellular localization will be investigated in the context of viral infection using the epitope tagged virus. The location of each protein alone will be first examined. It is conceivable that co-expression of the viral and host proteins might result in the relocalization of the host protein to a new intracellular compartment. This would be consistent with the interpretation that the two proteins interacted either immediately before or during intracellular relocalization. Fluorescence resonance energy transfer (FRET), which exploits

the fact that energy can be transferred from a fluorophore donor in the excited state to a fluorophore acceptor when it is in close proximity (<100 Angstroms), can be adopted to investigate the direct association between two intracellular molecules. (Jares-Erijman and Jovin, 2006). Reciprocal immunoprecipitation will be employed to validate specific interactions. In addition, RNAi/siRNA can be used to inquire about the importance of a suspected interaction to viral replication, as exemplified by the study of Tsg101 depletion (Garrus et al., 2001). Tsg101 complete 'knockdown' was achieved and resulted in a significant reduction of viral particle release and infectivity, which could in turn, be rescued by re-introduction of the Tsg101 gene. Furthermore, immunoprecipitation using various truncated forms of the viral and cellular proteins can be employed to map the binding domains or sites in both proteins, which can also be confirmed by immunofluorescence microscopy. Finally, *trans*-dominant mutants might be generated for suspected host interacting factors and used much as RNAi/siRNA is to abrogate the function of the cognate wild type factor during the process of infection or in virion-producing cells.

It is possible that a single viral protein can bind to different cellular proteins at different times and/or locations during various steps of the HIV-1 infectious cycle, especially for a protein like HIV-1 integrase that is proposed to be multi-functional. To be more comprehensive in our study, our approach can be used to investigate interactions at particular stage during viral life cycle by synchronizing infections. We believe our approach will allow us to observe rare, transient

interactions that would only become apparent in synchronized HIV-infected cultures at various points of viral replication or stages of the cell cycle. A synchronized infection can be achieved by infecting target cells with a replication competent tagged virus for 0.5 hours at 15^{o}C. The cells will then be washed extensively to remove extra viruses attached to the cells or in the medium and placed at 37^{o}C to initiate infection. Therefore, viruses will enter the cells at approximately the same time point starting the replication cycle. Synchronously infected cells will be harvested and assayed at selected time points (i.e. 2, 8, 12 or 24 hours) post-infection. The tagged viral protein(s) recovered at various time points within a single-cycle infection has the potential to retrieve entire complexes at particular stages. These types of studies will not only reveal the dynamic extent of HIV-1 replication and its cellular cofactors, but also may increase the chance of detecting any transient integrase-cellular complexes normally existing at low concentration since all of the viruses will be at the same step at the same time.

In addition to identifying regions within the proteins that can tolerate insertions without adverse effects on viral growth, the transposon linker scanning libraries can also be used for structure/function characterization of HIV-1 viral proteins. If the protein of interest is saturated with a library of random insertions, the distribution of the insertions can reveal important information with regard to different regions of the protein. For example, regions of a viral protein less amenable to linker insertion (i.e., regions in which few viable tagged derivatives

can be obtained) indicates that this region of the protein is critical to protein structure or function, while insertional “hot spots” suggests that those regions are highly flexible sequences or possibly surface accessible. Predictably, organisms with small genomes and regions of overlapping reading frames will have very stringent requirements with regard to insertional tolerance. Consistent with this interpretation, only one or two places within the three HIV-1 proteins we initially targeted could accommodate a five amino-acid insertion. However, this survey although comprehensive for the presence of diverse insertional points within the specific viral protein, was not exhaustive with respect to the independent and reiterative selections for viable tagged mutant viruses. It is not inconceivable that if this system was pushed further, new insights about viral protein structure/function relationships might emerge.

Epilogue

Studies presented here provide an innovative strategy to elucidate of the complex interactions between cellular and viral proteins during HIV infection, and can be extended to other areas of research. It should be a great asset not only for providing insight in HIV biology but also for discovery of unexpected targets for its antiviral intervention.

The strategy developed here can be used to generate replication competent viruses with an epitope tag insertion in any viral coding region. As a result, it is a powerful molecular tool that can lead to the identification of a variety of host

factors (and their complexes) interacting specifically with the tagged viral proteins during an infectious viral cycle. This systematic approach could possibly identify networks of cellular proteins that are involved in viral replication. The utility of tagged viruses extends beyond the scope of cellular protein interaction studies. For instance, the viruses with specific biochemical or immunologic tags can be used to examine the intracellular localization of viral proteins with much greater sensitivity than in the past. This will allow the visualization of tagged viral proteins during infection. Indeed, HIV-1 IN has been labeled with a small tetracysteine tag, which preserved infectivity of the virus while allowing it to be labeled with the bis-arsenical fluorescein derivative, FlAsH. This labeling allowed image processing of both intracytoplasmic and intranuclear HIV-1 complexes in three dimensions over time (Arhel et al., 2006). Insertion of a tag for visualization will provide new insights into the various movements of HIV-1 complexes within infected cells and provides the basis for the study of the dynamic virus-host cell interactions that occur during infection.

Appendix 1: HIV-1 Integrase Interacts with RNA Helicase A in vitro

A 1.1 Introduction

HIV-1 integrase (IN) is a viral protein that displays pivotal roles in virus replication and pathogenesis. It is the key player in viral cDNA integration into the host chromosome (Coffin et al., 1997). In addition, this viral protein has also been shown to contribute to other steps during the early stages of viral replication, including reverse transcription (Wu et al., 1999) and viral DNA nuclear import (Ao et al., 2007). Several cellular binding partners for IN have been reported, including SNF5/INI1 (Kalpana et al., 1994), LEDGF (Cherepanov et al., 2003; Kalpana et al., 1994), and p300 (Cereseto et al., 2005). The identification of these cellular factors has provided valuable insights into IN function.

In an attempt to search for cellular factors that interact with HIV-1 IN, we used *in vitro* affinity purification coupled with mass spectrometry. We show that IN can specifically recover RNA Helicase A (RHA) in various cells using a variety of IN-derived probes. In addition, the subcellular localization of both proteins was investigated.

RNA helicase A (RHA, gi|100913206) is a member of the DE*X*H-box (where *X* can be any amino acid) family of proteins. It is also termed nuclear DNA helicase

II (NDHII) or DHX9 (Rocak and Linder, 2004). RHA contains a helicase core domain consisting of seven motifs that are conserved in all RNA helicases and two copies of a double-stranded RNA binding domain in the N-terminal region (Zhang and Grosse, 1997). RHA is a nuclear protein that shuttles between the nucleus and the cytoplasm with the assistance of a bidirectional nuclear transport domain consisting of 110 amino acids within the C-terminus (Tang et al., 1999). Its function is subject to regulation by arginine methylation, a modification catalyzed by protein-arginine methyltransferase 1 (PRMT1) (Smith et al., 2004). RHA is able to unwind double-stranded RNA or DNA with the energy derived from the hydrolysis of NTPs by virtue of its NTPase activity (Lee and Hurwitz, 1992). This property enables RHA to participate in multiple cellular processes from RNA transcription to RNA processing to RNA nuclear export.

In light of its multiple functions in cellular RNA metabolism, it is reasonable to question if RNA helicases, or RHA in particular, regulates the activity of HIV-1 RNA. Indeed, RHA has shown to stimulate transcription of HIV-1 RNA (Fujii et al., 2001). The underlying mechanism appears to involve the binding of RHA to the transactivation response element (TAR) located at the very 5' end of all HIV-1 RNAs. This interaction is mediated by the double-stranded RNA binding domains within the N-terminus of RHA and is reminiscent of the interaction between TAR and the TAR-binding protein (Gatignol et al., 1991). It is perceived that, by virtue of its ability to bind TAR, RHA either modulates the transactivation activity of Tat or can indirectly influence RNA polymerase II in initiation/elongation from the viral

long terminal repeat (LTR). In addition, RHA also seems to enhance the Rev-Rev response element (RRE) control of late viral gene expression (Li et al., 1999). Since RHA itself exhibits weak interaction with the RRE and is thus unlikely to play a direct role in the export of the genomic viral RNA, RHA may promote the release of these unspliced RNAs from spliceosome.

In addition to its involvement in HIV-1 gene expression, a recent finding has revealed that RHA contributes to HIV-1 particle assembly and reverse transcription. RHA is shown to associate with constituents of the HIV-1 Gag protein and is packaged into virions in an RNA-dependent manner (Roy et al., 2006). It is demonstrated that siRNA mediated RHA-knockdown produces significantly less infectious viral particles. Thus it is conceivable that RHA gains access to the viral assembly machinery simply by binding to HIV-1 RNA. Interestingly, defective particles with no or diminished levels of RHA also showed reduced virion-associated endogenous reverse transcriptase activity. In this respect, RHA might assist HIV-1 reverse transcriptase to more efficiently copy viral genomic RNA by unwinding its secondary structure or promoting the interaction of this RNA with the nucleocapsid protein in order to assemble into an active reverse transcription complex.

A 1.2 Results

A 1.2.1 Precipitation of RHA from Various Immunoprecipitations with Integrase or Its SH3 Domain

To investigate cellular factors that interact with IN, various DNA constructs were created to fuse integrase or its C-terminal SH3 domain (amino acids 220-270) to high affinity tags such as the 3xFLAG epitope or EGFP-MBP. The 3xFLAG tag was inserted at IN residue 280 instead of the N- or C-terminal end. Since 3xFLAG tag insertion at this particular position results in a replication competent virus (see above), it is likely that ectopic expression of IN with the 3xFLAG-tag incorporated at residue 280 maintains proper IN structural/functional relationships and as such, is likely an excellent substrate to probe natural host interactions encountered by this protein in the cell. The 3xFLAG-tagged IN was expressed transiently in transfected 293T cells. Following immunoaffinity purification, SDS/PAGE and Coomassie blue staining was used to visualize recovered proteins. Individual protein bands were excised and subjected to mass spectrometry analysis, as shown in Figure A1.1. RHA was identified in the band corresponding to the molecular weight RHA of 140 kDa (Nakajima et al., 1997).

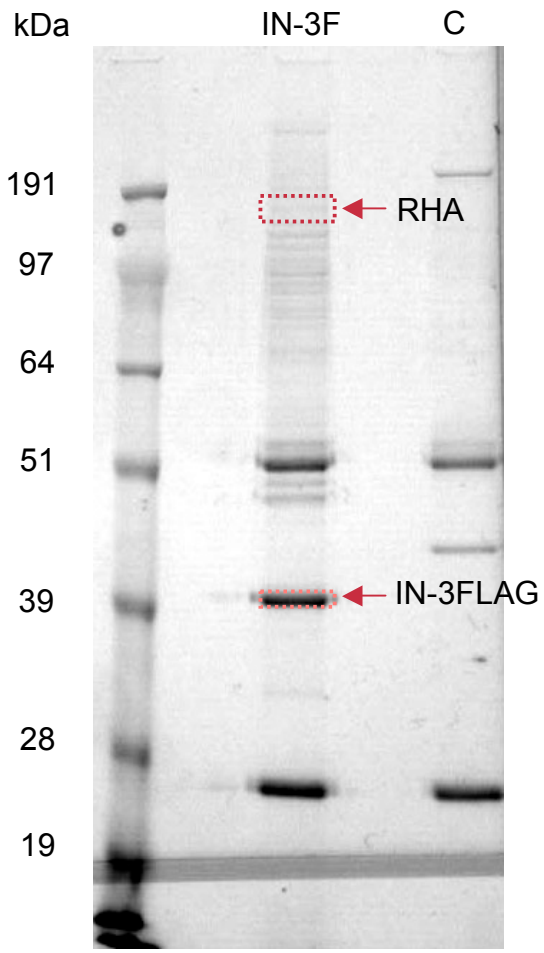


Figure A1.1 Immunoaffinity purification of protein complexes using 3xFLAG-tagged IN and visualized by Coomassie staining. RHA and IN, boxed and indicated with arrows were identified by mass spectrometry. Control (C) was transfected with pcDNA3.1.

Previous reports have shown that the integrase C-terminal domain (IN CTD) is important for various functions of integrase (Lu et al., 2005; Lutzke and Plasterk, 1998). Research in our lab has focused on the nuclear localization property associated with the IN-CTD, which has been mapped to about 45-50 amino acids that in three dimension elaborates an SH3 domain (220-270). We have shown that when IN-CTD is fused with EGFP-MBP (67 kDa) it resides in the nucleus, while a triple Lys-to-Ala substitution mutation at positions 262, 263 and 266 results in its cytoplasmic localization. To search for cellular proteins that bind to this particular region, EGFP-MBP was fused to the IN-SH3 subdomain. The triple mutant fusion protein EGFP-MBP-IN($SH3^{K(262/3/6)A}$) was constructed as well and used as a control. Figure A1.2 shows the immunoisolation of the fusion protein using α-GFP antibody. Transfection efficiency was determined by the green fluorescence associated with EGFP expression. In this experiment more than 50% of the 293T cells express the EGFP fusion protein. The majority of the IN-SH3 fusion was observed in the nucleus while the triple mutant fusion was restricted to the cytoplasmic compartment (Figure A1.2). Endogenous RHA was identified by mass spectrometry in the sample isolated from the wild type IN SH3 domain fusion protein in the region of 150 kDa. The entire sample and control lanes were subjected to MS analysis. RHA was absent in control lane of EGFP-MBP-IN($SH3^{K(262/3/6)A}$).

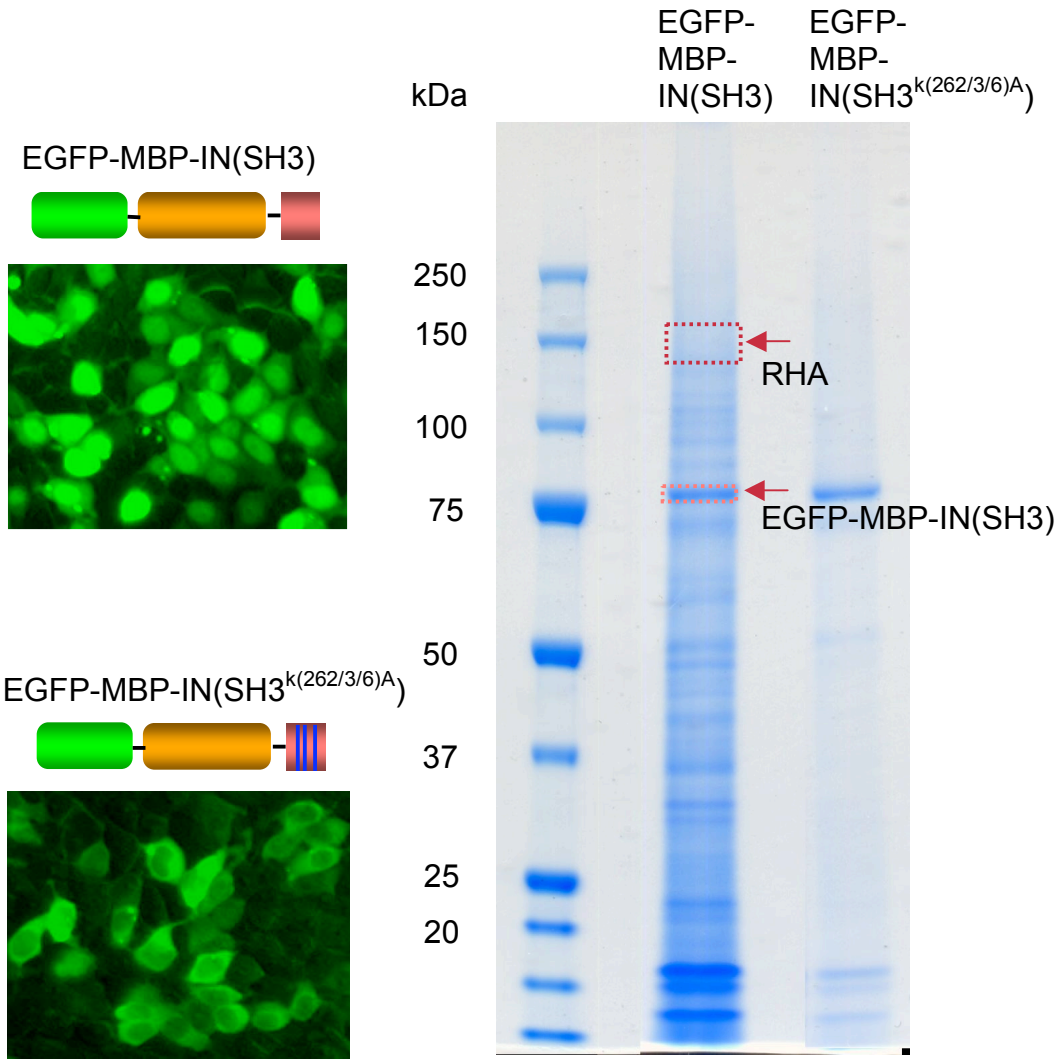


Figure A1.2 Immunoaffinity purification of EGFP-MBP-IN(SH3) and its associated proteins. The expression of green fluorescence and the subcellular localization are shown on the left. RHA and IN, boxed and indicated with arrows were identified by mass spectrometry. (in collaboration with Dr. Ileana Cristea in the Laboratory of Dr. Brian Chait).

Similarly, RHA was specifically immunoprecipitated from a CEM cell lysate using amylose magnetic beads bound to recombinant MBP-IN(SH3) fusion protein expressed and purified from bacteria (Figure A1.3). Here, an unrelated SH3 domain, Sso7d (Baumann et al., 1994), was fused with MBP to create MBP-Sso7 and used as a structurally-similar control. Unfused MBP was also used as a negative control. RHA was not identified from control bands cut from the gel in the region of interest. This system differs from the one described above for transfected mammalian cells because the fusion protein was first produced in *E. coli* and then bound to amylose magnetic resins before its incubation with a target T cell lysate and therefore provides evidence that IN may interact with RHA in an unmodified form.

A 1.2.2 Validation of the RHA-IN Interaction by Immunoblotting

To verify the interaction between RHA and IN, N-terminal 3xHA-tagged RHA and 3xFLAG-tagged IN^{280} were co-transfected in 293T cells. As controls, cells were transfected with 3xHA-RHA or IN-3xFLAG alone. Immunoprecipitation was performed with α-FLAG. As shown in Figure A1.4, the presence of 3xHA-RHA in the affinity purification was detected when the membrane was probed with α-HA. There were no signals corresponding to 3xHA-RHA in the control lanes where tagged RHA or IN was separately expressed. These results indicate that 3xHA-RHA can be specifically immunoprecipitated with IN-3xFLAG. However, reciprocal precipitation using α-HA conjugated beads failed to pull out 3xFLAG tagged integrase under the conditions used.

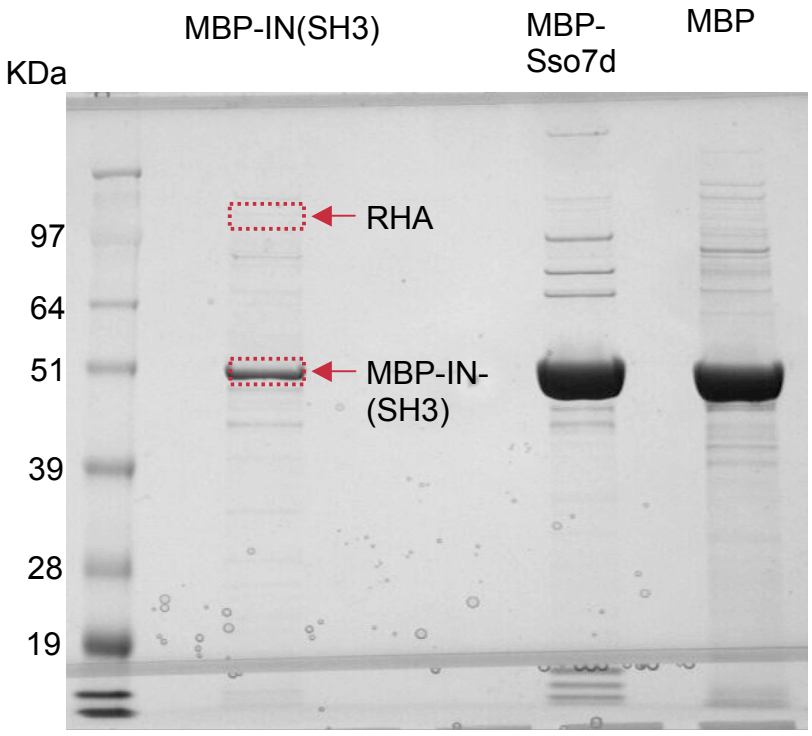


Figure A1.3 Immunoaffinity purification of MBP-IN(SH3) and its associated proteins using a bacteria-mammalian system as described in the text. RHA and MBP-IN(SH3) were confirmed by MS analysis and indicated with arrows. (in collaboration with Dr. Ileana Cristea in the Laboratory of Dr. Brian Chait).

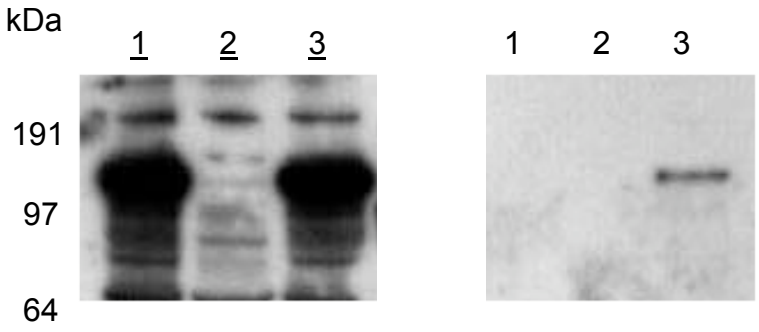


Figure A1.4 Immunoprecipitation of 3xRHA with IN-3xFLAG. Cells were incubated with α-FLAG and probed with α-HA. Elution lanes are on the right. (1) 3xHA-RHA alone, (2) IN-3xFLAG alone, (3) 3xHA-RHA + IN-3xFLAG. A portion of the input total lysates is probed on the left.

A 1.2.3 Subcellular Localization of RHA and Integrase

RHA is a nuclear protein and shuttles between the nucleus and the cytoplasm (Tang et al., 1999). Expression of HIV-1 integrase in mammalian cells is mainly localized to the nucleus. To investigate to what extent RHA and IN co-localize within cells such that they can interact, 293T cells were co-transfected with EGFP fused to the N-terminal of IN (EGFP-IN) and 3xFLAG tagged RHA (3xFLAG-RHA). Indirect immunofluorescent staining was performed using α-FLAG. As shown in Figure A1.5, 3xFLAG-RHA was mostly observed in the nucleus with a small, but detectable amount in the cytoplasm. Interestingly, RHA was observed to be slightly more concentrated at the nuclear envelope (Figure A1.5 B and C). This is consistent with the report that RHA is present in a highly enriched NPC fraction (Gatfield et al., 2001). Similar to RHA, both EGFP-IN and EGFP-MBP-IN(SH3) were mainly localized in the nucleus with trace amounts in the cytoplasm. When the fluorescent signals for RHA and EGFP-IN (or EGFP-MBP-INSH3) were merged, we observed that wild type integrase and RHA were both localized to the nucleus but at this level of discrimination, no evidence for co-subnuclear localization could be established. On the other hand, the nuclear localization triple mutant fusion protein EGFP-IN$^{K(262/3/6)A}$ was excluded from the nucleus. It did not change the distribution of RHA, consistent with our finding that the IN triple substitution mutant does not interact with RHA.

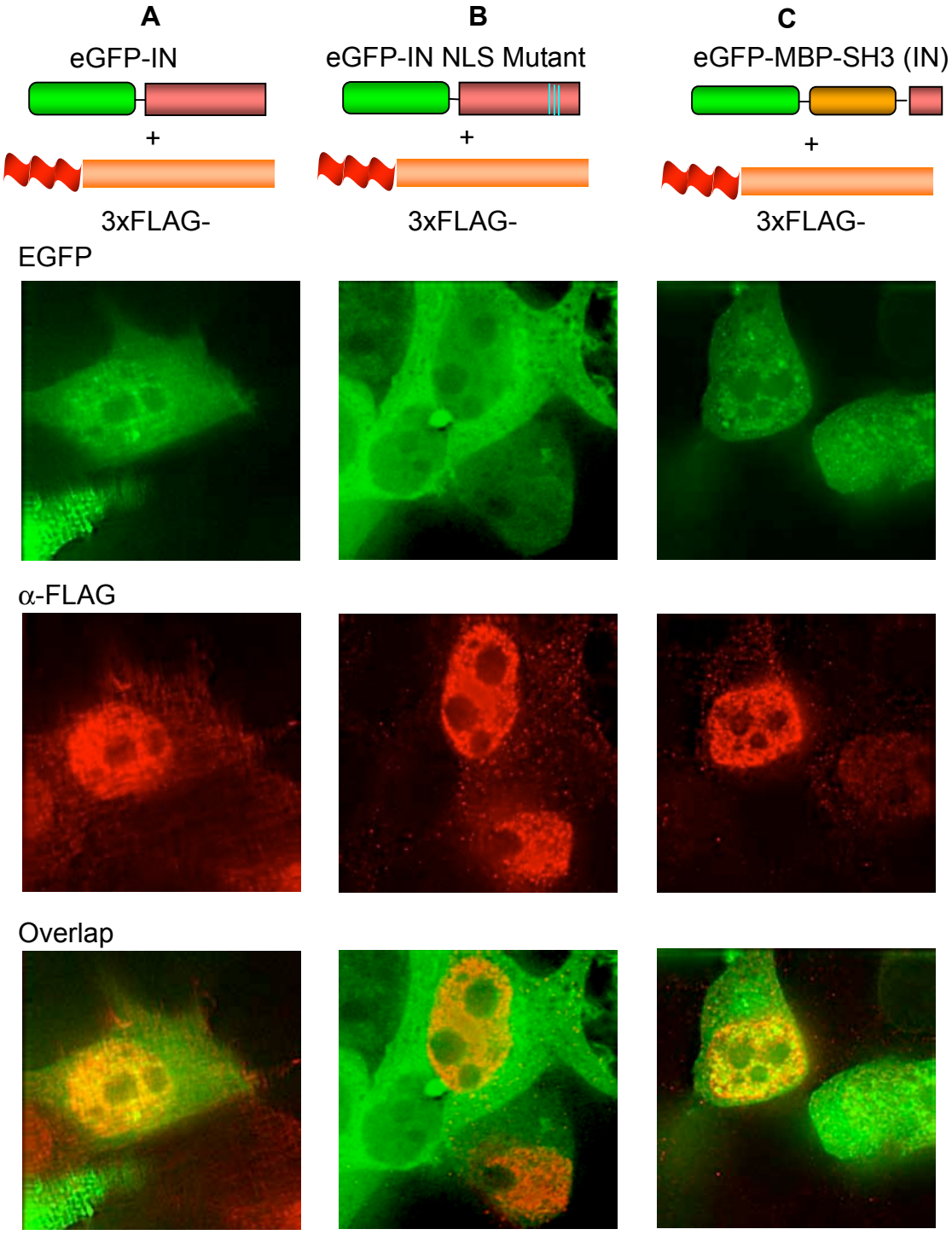


Figure A1.5 Subcellular localization of 3xFLAG-RHA and various EGFP-IN fusions in 293T cells.

A 1.3 Discussion

In this study, we provide evidence that RHA associates with HIV-1 integrase. Integrase is an important viral enzyme and a component of the machinery that provides nuclear localization, integration and coordinates DNA repair for the virus. Taking the results obtained from other laboratories into consideration, we speculate that RHA might be involved in multiple steps of viral life cycle through its interaction with IN. It might exert an effect through modeling the secondary structures of viral RNA or rearranging RNA-protein interactions. Considering the sophisticated protein-protein and protein-RNA interactions that take place during generation of virus particles, it is not surprising that the virus utilizes the RNA binding ability of an RNA helicase, like RHA, to regulate these complex molecular events.

RHA is one of the few RNA helicases that have been reported to play a role in HIV-1 replication. It has been found to be incorporated into the HIV-1 virion and may participate in HIV-1 reverse transcription as well as particle assembly (Roy et al., 2006). It has also been reported to impact HIV-1 transcription through its interaction with TAR (Fujii et al., 2001). HIV-1 particles that do not contain RHA show reduced virion endogenous RT activity (Lu et al., 2005; Wu et al., 1999). It is possible that RHA may directly assist viral RT in the coping of vRNA into cDNA through unwinding RNA secondary structures or regulation of the interactions RT makes with other factors to form a functional reverse transcription complex. Coincidentally, IN has been documented to affect RT activity as well. It would be

interesting to determine if RHA play a direct role in reverse transcription or indirectly influences RT activity through its interaction with IN. A summary of possible RHA and HIV-1 interaction nodes is summarized in Figure A1.6.

The function of RHA might be related to HIV-1 replication in other ways. The regulatory activity of RHA in RNA transcription is implicated by its presence within the RNA polymerase II holoenzyme complex. For example, RHA has been shown to bridge the interactions made between RNA polymerase II and transcription co-activators such as CREB-binding protein and BRAC1 (breast cancer-specific tumor suppressor protein 1) (Nakajima et al., 1997). RHA also directly interacts with the p65 subunit of NF-B and stimulates NF-B-mediated reporter gene expression (Tetsuka et al., 2004). With regard to RNA nuclear export, RHA directly binds to the constitutive transport element (CTE) RNA of type D retroviruses. In addition, together with the cellular factors Tap (Tip-associated protein) and HAP95 (helicase A-binding protein 95), RHA promotes the export of CTE-containing RNA molecules (Tang et al., 1997). Finally, RHA has been shown to interact with Ubc9 and leads to the SUMO-1 conjugation of RHA both *in vitro* and *in vivo* (Argasinska et al., 2004).

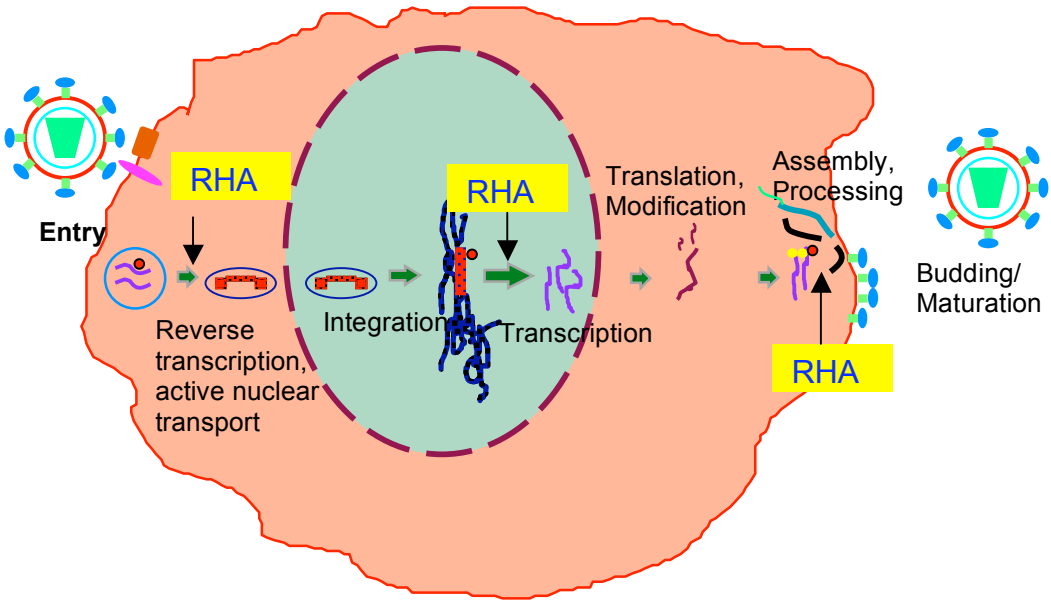


Figure A1.6 A Schematic representation of the possible involvement of RHA at the steps of reverse transcription, proviral transcription and in downstream post-transcriptional events of HIV-1.

Research in our lab has shown that HIV-1 IN interacts with the DNA repair protein Rad18 (Mulder et al., 2002) and part of the Rad6 DNA repair pathway is associated with modification of PCNA by ubiquitin (Hoege et al., 2002). Furthermore, Rad18 forms a tight complex with Rad6 both in yeast and vertebrate cells (Bailly et al., 1997; Tateishi et al., 2000) and it has been shown to be mono- and poly-ubiquitinated in a Rad6-dependent manner (Miyase et al., 2005). Rad6 is a member of the E2 Ub-ligase family (also named UBC2). Rad18, like many other E3 enzymes, contains a RING finger domain. In both yeast and vertebrate cells, Rad18 also contains the RING domain within its NH2-terminal region. This may suggest that Rad18 functions as an E3 recognin. It was subsequently found that Rad18 is involved in the monoubiquitination of PCNA in cells treated with DNA-damaging agents – PCNA monoubiquitination critical for polymerase switching in *trans*-lesion DNA synthesis (Hoege et al., 2002; Kannouche et al., 2004; Stelter and Ulrich, 2003). It would be interesting to see whether RHA is ubiquitinated and functions through the same or a similar pathway during HIV-1 infection via its interaction with IN.

It has been suggested that RHA initiates its interaction with HIV-1 genomic RNA within the nucleus and that this interaction persists throughout the subsequent events that occurs until the genomic RNA is incorporated into viral particles (Roy et al., 2006). To resolve whether there is direct interaction between IN and RHA or whether it is RNA dependent, the effects of RNase treatment to RHA immunoprecipitation with IN is being studied. Preliminary results showed that the

addition of RNase did not eliminate the interaction between RHA and IN. However, further experiments such as addition of an RNase inhibitor, RNA detection or binding experiment using purified proteins are necessary to confirm this observation.

Our experiments using IN SH3 domain fusion proteins shows that the SH3 is sufficient for the binding of IN to RHA. Although RHA was recovered with full-length IN, we have so far been unable to show reciprocal immunoprecipitation using N-terminal-tagged RHA. It is possible that the N-terminal tagging in RHA interferes with the putative RHA-IN interaction. Immunoprecipitation using C-terminal tagged RHA will help address this potential problem. Although subcellular localization studies demonstrate that both proteins are predominantly within the nucleus, we are unable to reach a conclusion concerning the potential co-localization between RHA and IN within this cellular compartment. Fluorescent resonance energy transfer (FRET) analysis might help to address this question. Detailed MS or MS/MS analysis of immunoaffinity isolation of IN-3xFLAG tagged viral infection is underway to investigate if RHA is present in the elution. This would be a further indication of an RHA-IN association that is active during active viral infection.

In summary, our study demonstrates the association of RHA with ectopically-expressed HIV-1 IN by mass spectrometry analysis and verified by immunoprecipitation. Further experiments, including reciprocal

immunoprecipitation, FRET microscopy and mapping of the binding domain are necessary to authenticate this interaction.

Appendix 2: Investigation of HIV-1 Integrase Post-Translational Modifications

A 2.1 Introduction

Post-translational modifications (PTMs) are covalent modifications that change the properties of a protein by proteolytic cleavage or by addition of a modifying group to one or more amino acids. They modulate the activity of most eukaryote proteins by defining its active state, localization, turnover and interactions with other proteins. Examples of the biological effects of protein modifications include phosphorylation in signal transduction pathways (Cohen, 2000), ubiquitination-dependent proteolysis (Tyers and Jorgensen, 2000), transcriptional activation via acetylation (Wang et al., 2005), attachment of fatty acids for membrane anchoring and association (Resh, 2006) and glycosylation can affect protein half-life (Sinclair and Elliott, 2005). Analysis of these modifications presents formidable challenges but their determination can generate indispensable insight into biological function. To map modification sites in molecular detail, biochemical characterization (Gesellchen et al., 2006), site directed mutagenesis (Ragolia et al., 2007) and novel mass spectrometric peptide sequencing and analytic technologies have been successfully applied (Larsen et al., 2006; Mann and Jensen, 2003).

Post-translational modification is of great interest for the study of viruses. Pathogens have coevolved with their hosts and acquired the ability to optimize their replication by taking advantage of cellular signaling networks (Manes et al., 2003). This strategy of survival becomes extremely important for small parasites such as viruses. The post-translational state of many viral proteins might have evolved to intercept, disrupt, mimic or usurp numerous signaling pathways in their hosts (Kahn et al., 2002).

HIV-1 IN has been shown to undergo post-translational modification by acetylation (Cereseto et al., 2005). HIV-1 IN is a substrate for histone acetyltransferase (HAT)-mediated, p300 posttranslational acetylation both *in vitro* and *in vivo.* Interestingly, p300 modification targets lysine residues within the C-terminal domain of integrase (IN-CTD). The IN-CTD is known to possess intrinsic DNA binding activity *in vitro* (Engelman et al., 1994; Gao et al., 2001; Lutzke and Plasterk, 1998). The ability of many proteins to bind DNA is affected by p300 acetylation (Gu and Roeder, 1997; Wang et al., 2005). It is of great interest to see if posttranslational acetylation by p300 in some way regulates the function(s) of HIV-1 integrase. To confirm the above observations and to characterize the reported acetylation mutant, we mutagenized an R7/3 proviral clone at the IN codons for lysine (K) residues 264, 266 and 273. In this mutant (3R) the lysines, previously identified as the primary targets for acetylation, are substituted with arginine (R), In the Cereseto report, this particular K>R triple substitution mutant is defective for replication (Cereseto et al., 2005). However, when these specific

substitutions were incorporated into R7/3, our standard HIV-1 wild type proviral clone, the R7/3 3R mutant virus replicated with kinetics similar to wild type. This discrepancy might be explained by the Cereseto group's use of an acidic epitope tag (1xFLAG) placed at the extreme C-terminus of integrase (position 288/288) near the sites targeted for acetylation. We showed that while the combination of the epitope tag with the RRR substitution results in a replication-defective phenotype , both the 3R mutant and 1xFLAG viruses are viable when in single combination..

Phosphorylation of integrase has been reported for the integration protein of Rous sarcoma virus (RSV), an avian retrovirus (Horton et al., 1991). A phosphorylation site has been identified in the C-terminus of RSV IN and is thought to regulate correct proteolytic processing of RSV IN in the producer cell. Research in our lab has also shown that HIV-1 IN can also be phosphorylated *in vitro* by Casein Kinase II (CKII). CKII is a protein Ser/Thr kinase complex that forms a heterotetramer consisting of two catalytic (α and/or α') subunits and two noncatalytic (β) subunits (Litchfield, 2003). CKII is essential for viability in yeast (Glover, 1998; Kikkawa et al., 1992), which may, in part, protect cellular proteins from caspase proteolysis. There is also mounting evidence indicating that CK2 is a component of regulatory protein kinase networks involved in various aspects of transformation and cancer (Fraser et al., 2000; Landesman-Bollag et al., 2001). CKII has also been implicated in viral biology. In one study, a CKII site has been found to flank the nuclear targeting signal of Human cytomegalovirus (HCMV)

and enhance nuclear import of the HCMV ppUL44 protein (Alvisi et al., 2005). With regard to HIV-1, CKII has been shown to phosphorylate HIV-1 MA (Swingler et al., 1997), Rev (Meggio et al., 2001) and Vpu (Friborg et al., 1995).

Although four conserved CKII consensus target sites exist in HIV-1 IN, our study here identifies residue S283 as the key site of CKII phosphorylation *in vitro*. This result was obtained from the analyses of mutants containing either serine-to-alanine (S→A) or threonine-to-alanine (T→A) at the consensus CKII phosphorylation sites within IN. Single or combinatorial mutants were characterized by biochemical assay or mass spectrometry analysis. Among the predicted sites, the serine at residue 283 within the IN CTD is the primary residue targeted for CKII phosphorylation *in* vitro. Characterization of viruses incorporating the S283A mutation is described below.

Ubiquitination covalently attaches a single ubiquitin molecule or multi-ubiquitin chain to target proteins, giving rise to mono- or poly-ubiquitinated proteins, respectively. Most polyubiquitinated proteins are subjected to proteolysis by the proteasome. In contrast, monoubiquitination has been associated with proteasome-independent cellular functions including endocytosis, virus budding, chromatin remodeling and nuclear protein export to the cytoplasm (Hicke, 2001; Li et al., 2003). During the process of ubiquitination, the ubiquitin molecule is activated through its linkage with the ubiquitin-activating enzyme (E1). Activated ubiquitin is then transferred to one of the ubiquitin-conjugating enzymes (E2s).

Ubiquitin is finally conjugated directly to the target protein usually with the help of an ubiquitin ligase (E3) (Bonifacino and Weissman, 1998; Hershko and Ciechanover, 1998; Pickart, 2001).

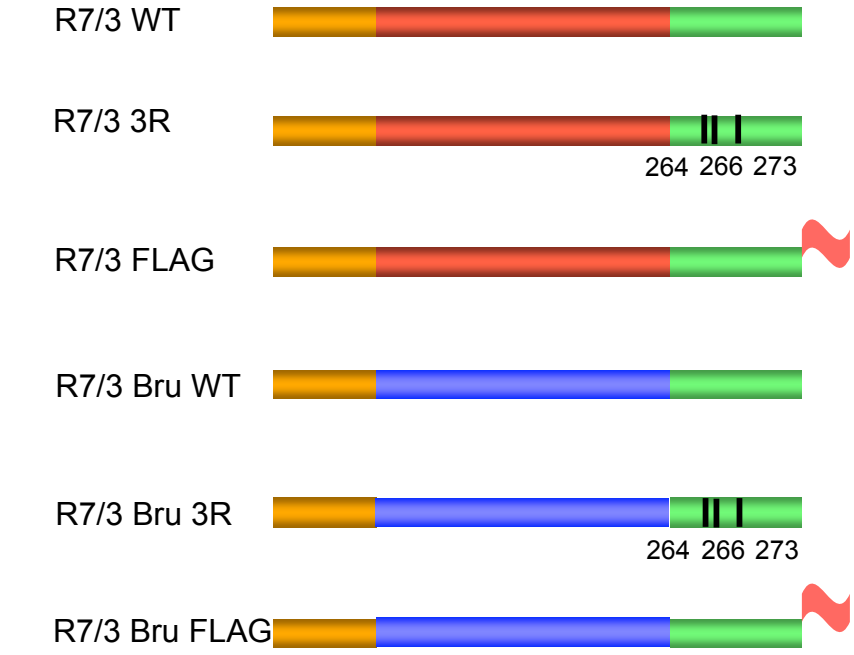
HIV-1 IN is a substrate for the N-end rule (Mulder and Muesing, 2000)—a distinct ubiquitin-proteasome pathway that defines the half-life of target proteins according to the identity of their N-terminal residue (Varshavsky, 1996). Authentic IN is the proteolytic product of the Gag-Pol polyprotein and specifies an N-terminal phenylalanine (Phe or F), which renders it particularly prone to rapid destruction by N-end rule proteasomal degradation when it is expressed in the absence of other viral proteins. In general, degradation of proteins by the proteasome occurs after the substrate has been covalently modified by linkage of a multi-ubiquitin chain, marking the protein for destruction. Rad18, a DNA repairing protein and an E3 ubiquitin ligase (Tasaki et al., 2005) and has been identified as an interacting partner of HIV-1 IN (Mulder et al., 2002). Paradoxically, however, the overexpression of Rad18 results in a corresponding increase in the stabilization of F-IN (Mulder et al., 2002). Here, we sought to biochemically and biophysically characterize IN ubiquitination *in vivo* (especially mono-ubiquitination of IN) by immunoprecipitation analysis and mass spectrometry. Using a panel of mutant Ub clones, our results indicate that IN is both monoubiquitinated and polyubiquitinated in cells.

A 2.2 Results

A 2.2.1 Characterization of an HIV-1 IN Acetylation Mutant

The previously reported (Cereseto et al., 2005) acetylation mutant with Lys to Arg substitution at residues 264, 266 and 273 (3R) was constructed in the context of the proviral clone Bru, a proviral clone with a genetic pedigree convergent to that of R7/3 (Wain-Hobson et al., 1991). To be consistent with the original finding, the 3R mutant was constructed in both Bru and R7/3 backgrounds. The different proviral constructs used in our experiments are illustrated in Figure A2.1.

Replication-Competent



Replication-Defective

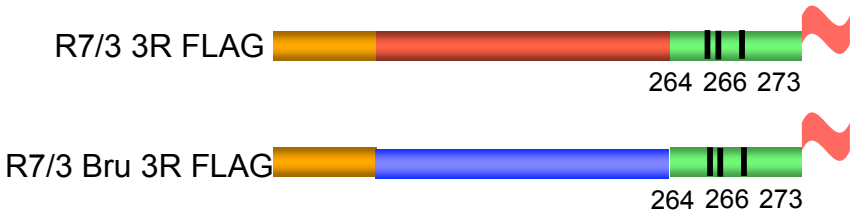
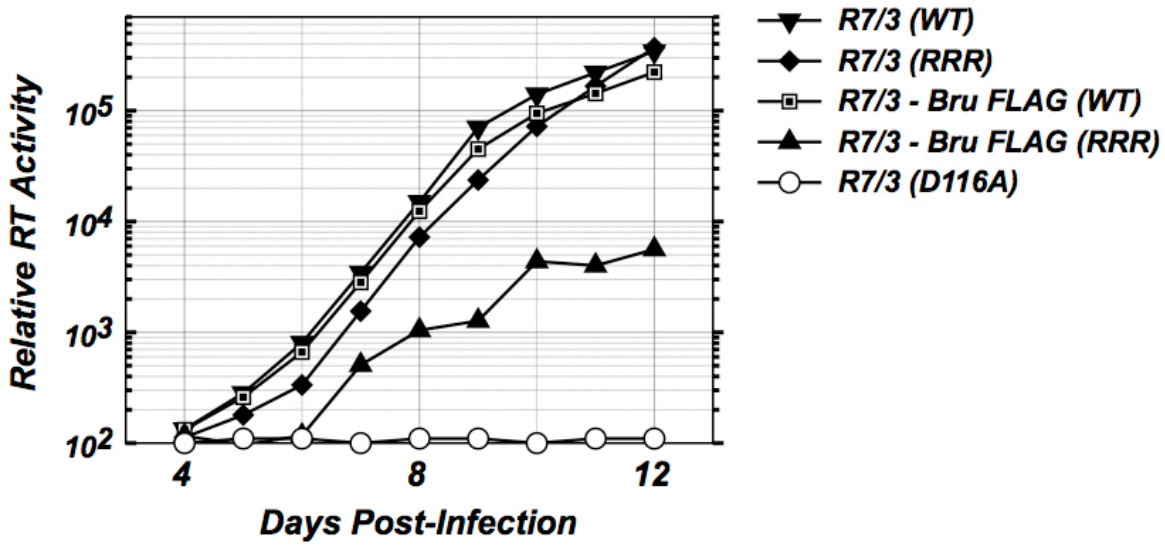


Figure A2.1 Summary of the integrase genotypes of the mutant viruses and their replication statuses in CEM cells. Viral infections were followed over the course of 12 days and scored for replication by syncytia formation and p24 or RT production.

Using viral growth curves generated in a manner similar to the original report, viral titers were determined by the activity of exogenous reverse transcriptase in the culture. As shown in Figure A2.2(A), the presence or absence of the integrase 1xFLAG tag is critical in the context of the mutations described. Thus, while the tagged and untagged wild type viruses are replication competent, addition of the 1xFLAG tag to the 3R substitution mutant results in viruses incapable of growth in culture. In contrast, an untagged 3R mutant replicates well. In an independent viral growth assay measuring the p24 content, similar viral growth curves were produced [Figure A2.2(B)]. Again, the IN 3R mutant tagged at IN-288 with 1xFLAG was defective. Similar results were obtained in the indicator cell lines, M5.25 and HeLa P4R5 as shown in Figure A2.3. Finally, when we compared the relevant variables in all possible combination (wild type vs. 3R mutant, 1xFLAG-tagged vs. untagged and IN R7/3 vs. IN Bru) our observations were further strengthened (Figure A2.1). It is clear that a replication defect occurs when the epitope tag is present in combination with the closely spaced 3R mutation, regardless of the particular IN used, as the 3R mutation by itself, does not preclude viral replication through several cycles of infection in culture.

A



B

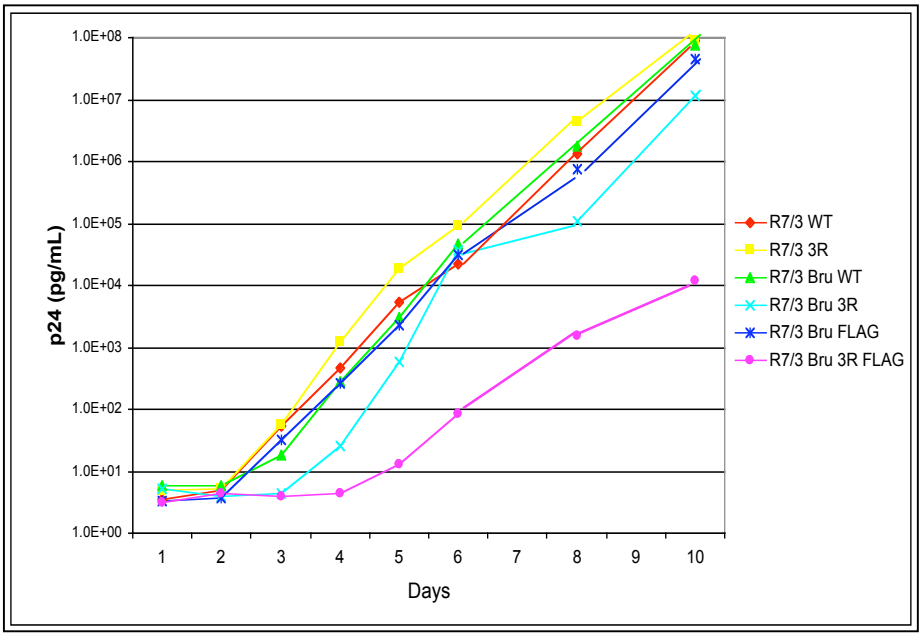


Figure A2.2 Replication potential of the WT and 3R mutant viruses in CEM cells. (**A**) Growth curves by RT assay. D116A is a catalytically inactive mutant and used as a replication-defective control. (**B**) Growth curves determined by p24 accumulation.

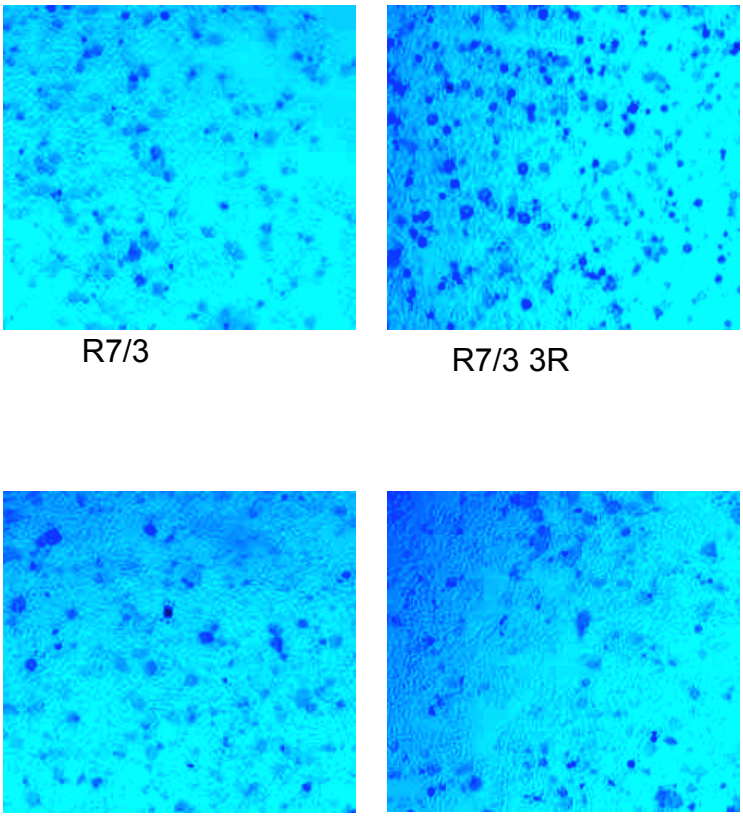


Figure A2.3 Comparison of the infection status of wild type and the K(264/266/273)R (3R mutant viruses) in HeLa P4R5 cells. The *in* gene from the Bru virus was cloned into R7/3 to produce the R7/3 Bru virus and its corresponding 3R mutant. Infected cells, shown in dark blue, were stained for β-galactosidase activity 2 days after infection.

A genetic assay using a dominant selectable marker, blastocidin D (*bsd*) was employed to measure the respective integration frequencies of the WT and 3R mutant viruses. Integration frequencies of greater than 15 to 20% of that of the wild type are required for sustained viral replication in CEM T cell culture (Wiskerchen and Muesing, 1995a, b). The untagged R7/3 3R mutant integrated with a frequency nearly equivalent to WT, whereas the 1xFLAG tagged mutant (R7/3 Bru 3R-1xFLAG) integrated with a frequency of about 12% compared to the tagged, wild type virus (R7/3 Bru-1xFLAG). This effect is not specific to the tagged Bru IN sequence *per se*, since the 1xFLAG-tagged mutant R7/3 (R7/3 3R-1xFLAG) when compared to R7/3-1xFLAG behaves in identical fashion.

To investigate replication dynamics of the 3R mutant in primary cell culture, especially after macrophage infection, an R5-tropic virus was used, R7/3:YU-2 in which the Env gene of YU-2 is recombined into R7/3 by replacement of the X4-tropic *env* gene. R7/3:YU-2 WT and its 3R mutant show no difference in their ability to infect macrophages. To see whether the 3R mutant preferentially utilizes a particular class of co-receptor, infections of WT and 3R in R7/3 (X4, CXCR4) or R7/3:YU2 (R5, CCR5) were studied in HeLa P4R5 cells. The results show that the 3R mutant is as infectious as WT (in some cases registering better in this assay) regardless of the co-receptor used for viral entry.

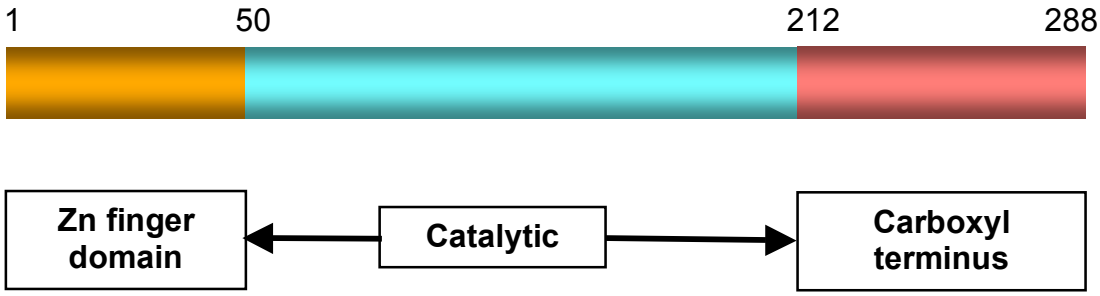
In attempt to verify the direct interaction between p300 and integrase *in vivo* (Cereseto et al., 2005), p300 tagged at its C-terminus (p300-HA) and a p300

expression clone with a 33 amino acid deletion mutant at its catalytic site (p300-HAT^{minus}) were generated and used for coimmunoprecipitation studies. Under the conditions used in our experiments, we failed to selectively immunoprecipitate endogenous p300, transfected p300-HA or p300-HAT^{minus} when 3xFLAG tagged IN was used as "bait". Reciprocal immunoprecipitations using p300-HA to probe potential p300 interactions with IN when examined in reverse also failed to recover 3xFLAG-tagged IN^{280} or an untagged full-length integrase after cotransfection with their respective expression plasmid.

A 2.2.2 Integrase is Phosphorylated by Casein Kinase II in vitro

Research in our laboratory has shown that recombinant IN is phosphorylated specifically by CKII among a panel of kinases tested *in vitro*. The consensus site for CKII-targeted phosphorylation is S/T-X-X-(E/D). As shown in Figure A2.4, there are four potential CKII phosphorylation sites conserved within the HIV-1 IN amino acid sequence: T66, T93, S195, S283. *In vitro* phosphorylation assays of recombinant His-tagged IN was performed using commercially obtained purified CKII (New England Biolabs) in the presence of ^{32}P-γ-labeled ATP. A titration experiment was performed to determine the linear range of kinase activity on its substrate *in vitro.* CKII was serially diluted and shown to phosphorylate the recombinant HIV-1 IN substrate using as little as 0.02 U/mL of CKII. Therefore, in comparison to other CKII substrates, IN is very active for modification by this kinase *in vitro* (Figure A2.5). To identify the primary site(s) of CKII modification, various Ser/The to Ala point mutations were constructed as shown in Table 7.

Wild-type recombinant His-tagged IN and its cognate mutants were purified and normalized for their protein content. Kinase reactions were performed using recombinant CKII. As shown in Figure A2.6, many of the mutations did not significantly reduce the level of IN phosphorylation. However, mutants of integrase that included the S283A substitution mutation were compromised for CKII phosphorylation. The S283A single mutant gave the weakest signal for phosphorylation due to slight less integrase in the particular sample. This indicates that the primary site for phosphorylation *in vitro* is at serine residue 283. Mutational analysis shows that phosphorylation can be reduced to 10% that of WT when alanine is substituted for serine at residue 283; the other potential phosphorylation sites in integrase are less reactive to this kinase.



Zn finger domain (1-49)

FLDGIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQ

LKGEA

Catalytic core (50-212)

MHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETG

QETAYFLLKLAGRWPVKTIHTDNGSNFTSATVKAACWWAGIKQ

EFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFI

HNFKRKGGIGGYSAGERIVDIIATDIQTKE

Carboxyl terminus (213-288)

LQKQITKIQNFRVYYRDSRNPLWKGPAKLLWKGEGAVVIQDNSDI

KVVPRRKAKIIRDYGKQMAGDDCVASRQDED

Figure A2.4 Consensus CKII phosphorylation sites within HIV-1 IN (T66, T93, S195, S283) shown in italicized bold red. The relative placement and sizes of the IN domains are indicated at the top of the figure.

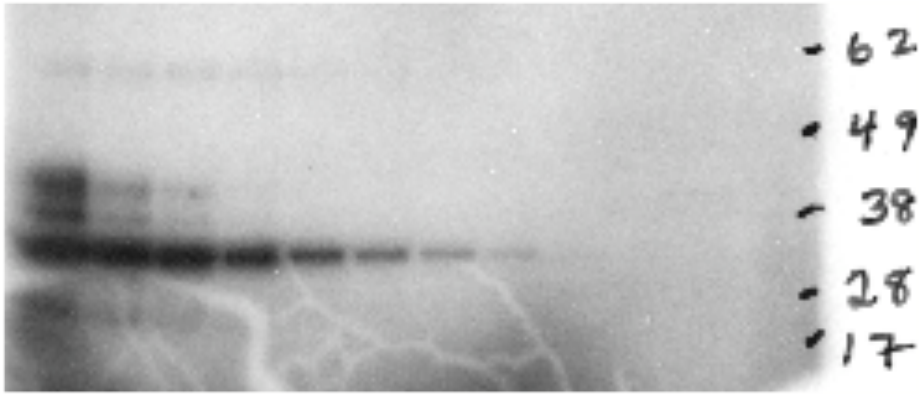


Figure A2.5 Titration of CKII activity on purified integrase. The CKII kinase reaction was performed using recombinant WT integrase in the presence of γ-^{32}P-ATP**.** CKII was serially diluted 1:3 starting from 50 U/mL (50mU/μL) and the extent of reaction visualized by autoradiography.

Table 7. His-tagged integrase alanine-substitution mutants altered at the targeted serine or threonine residues within consensus CKII sites.

Single Mutants	Double Mutants	Triple Mutants	Quadruple Mutant
T66A	T66/T93A	T66/T93/S195A	T66/T93/S195/S283A
T93A	T66/S195A	T66/T93/S283A	
S195A	T66/S283A	T66/S195/S283A	
S283A	T93/S195A	T93A/S195/S283A	
	T93/S283A		
	S195/S283A		

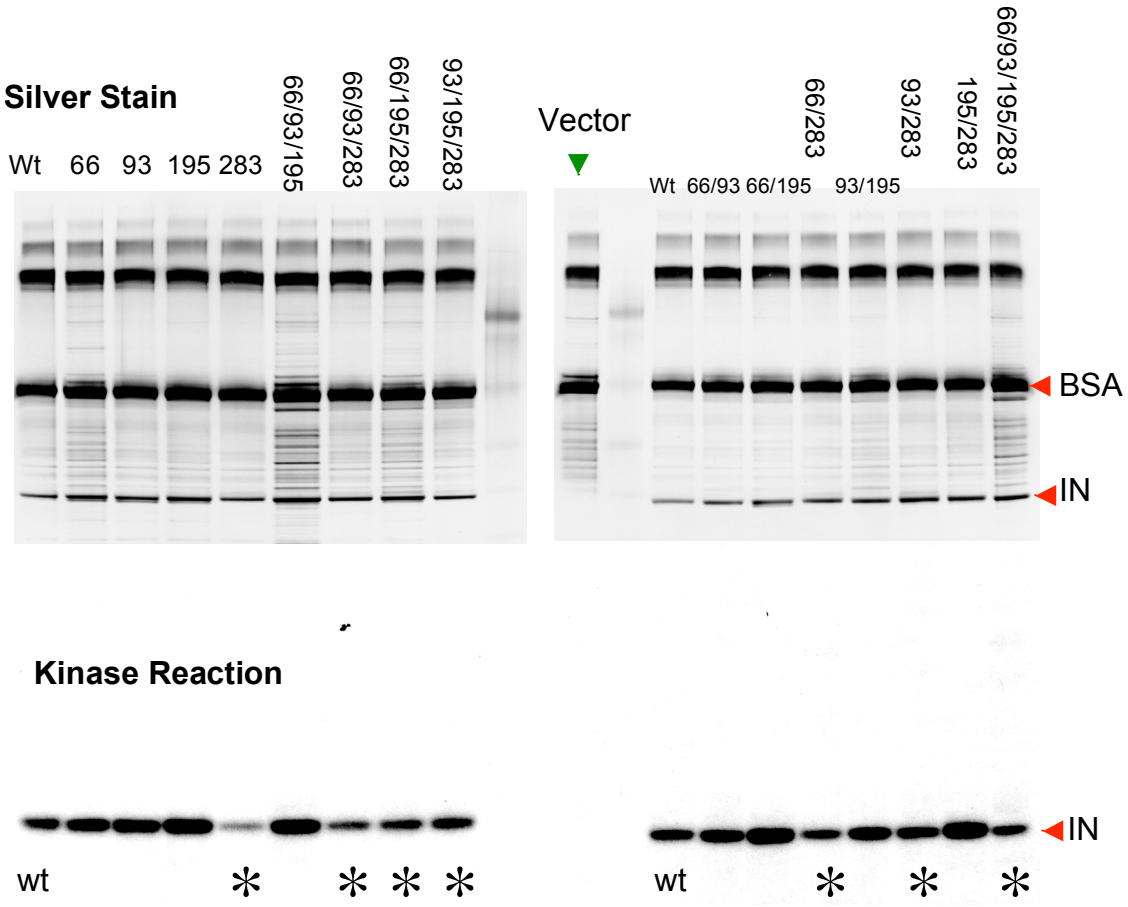


Figure A2.6 His-tagged mutants of integrase are compromised for CKII phosphorylation in S283A mutants (*). Samples were normalized for protein concentration. The silver stain of the same gel prior to autoradiography is shown at the top.

To map the IN domain responsible for CKII phosphorylation, IN and various truncated forms were fused with GST. A schematic illustration of the constructs and the Coomassie staining of the purified fusion proteins after normalization for their respective protein content are shown in Figures A2.7A and B. As a control, GST was used alone. Kinase reaction using CKII was performed on these GST fusion proteins as shown in Figure A2.7C. The full length IN(1-288) gave the strongest signal for phosphorylation when compared to IN(1-50) and IN(1-212) both of which were weakly phosphorylated. The IN-CTD (residues 212-288) is phosphorylated to a level comparable to that of the full-length IN(1-288). When the S283A mutation is incorporated into IN(1-288), the signal for phosphorylation is dramatically reduced. A similar result was obtained when the S283A mutation was recombined into the IN-CTD(212-288) fusion protein. Taken together, these results corroborate the earlier observations from the mutagenic survey of all potential CKII phosphorylation sites and indicate that in a variety of assays the serine residue at position 283 is the primary target of CKII activity, the IN-CTD being the main domain targeted for phosphorylation.

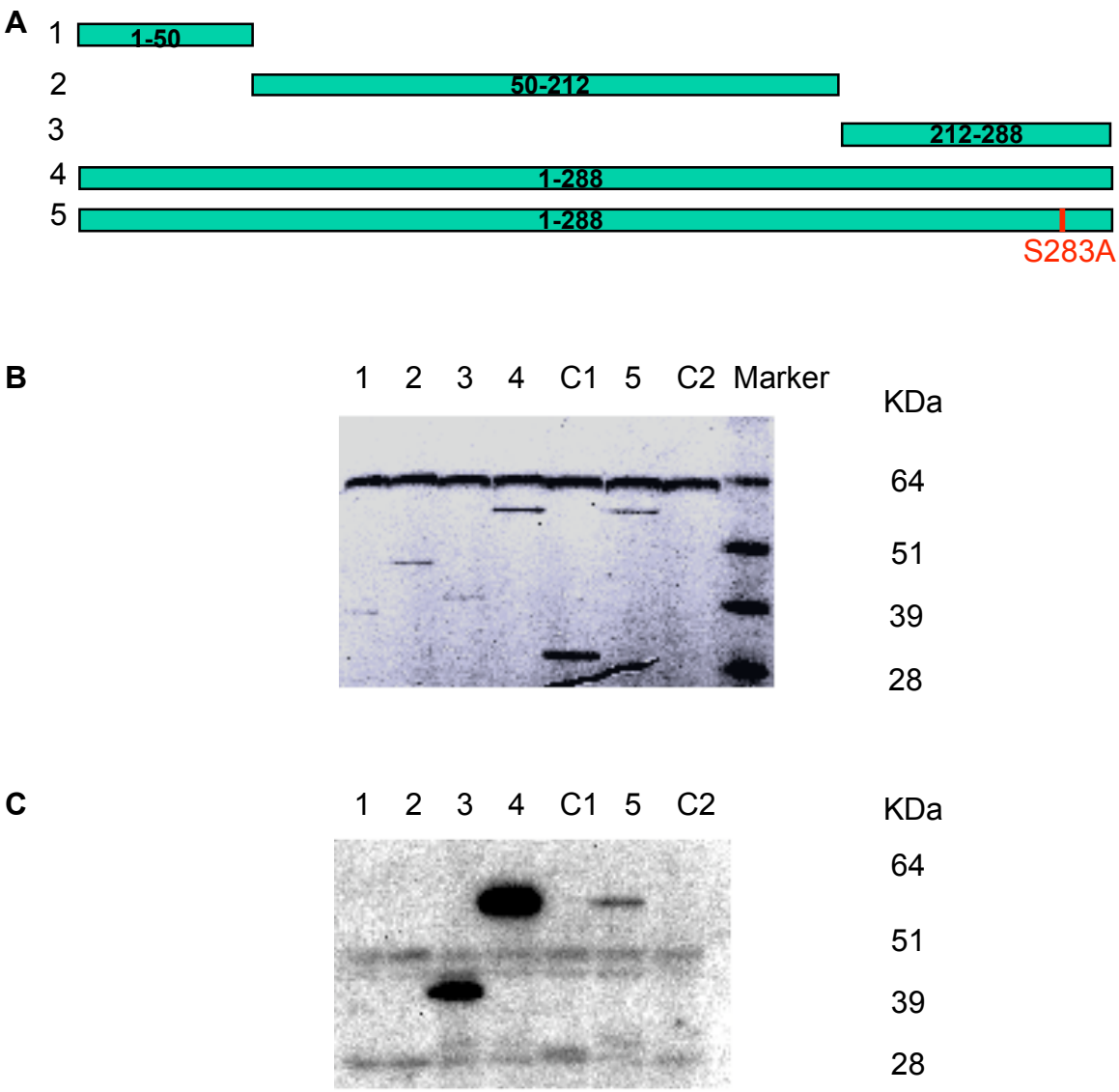


Figure A2.7 Mapping the integrase domain targeted for CKII phosphorylation using recombinant GST-IN, GST-IN(S283A) and GST-IN truncation fusion proteins as test substrates for CKII. (**A**) A schematic illustration of the GST constructs used. (**B**) Coomassie blue stain of the purified GST-IN fusion proteins used in the CKII reaction. (**C**) Autoradiography of the same gel in B. C1: GST alone. C2: no protein substrate.

In the viral context, when each of the four potential phosphorylation sites were mutated from S or T to alanine, T93A, S195A and S283A remained replication competent by syncytia formation. R7/3 T66A was found to be the only singly-substituted mutant virus that was defective as revealed by the lack of syncytia formation or detectable accumulation of p24. In contrast, the double or triple mutants that include S283A are observed to be viable. The subcellular localization of various phosphorylation mutants was also studied in the context of an EGFP-IN fusion. Most single and double mutations were nuclear localized except mutants that involve T66A, which are excluded from the nucleus.

Immunoprecipitation experiments were performed in the hope of probing a direct interaction between IN and CKII. Various strategies were used including reciprocal immunoprecipitation attempts using α-IN or α-CKII. Myc-CKIIβ and HA-CKIIα (or α'), 3xFLAG-IN and the 3xFLAG-IN(S283A) mutant were also tried using α-Myc, α-HA, or α-FLAG mAb immunoprecipitation. So far, we have been unable to detect any association between IN and CKII or any of its subunits in cells.

Since residual phosphorylation was observed with S283A mutant, MS/MS analysis was used in an attempt to identify other weakly kinased CKII sites in integrase. Wild-type recombinant integrase was purified and reacted *in vitro* with purified CKII. The band corresponding to IN was excised and submitted for MS analysis. Data analysis was based on a shift in mass (+80 Dalton) at the residue

targeted for phosphorylation. MS/MS analysis was able to show that S283 was indeed phosphorylated *in vitro*. However, the trypsin digestion failed to recover a stretch of peptide sequence from residue 50-120, where two of the potential sites T66 and S195 were located. The threonine residue at position 93, (T93) remained unphosphorylated in this experiment.

A 2.2.3 Integrase Ubiquitination

The first amino acid of authentic HIV-1 integrase produced during infection is phenylalanine (Phe), in contrast to recombinant HIV-1 IN that contains an N-terminal Methionine (Met) residue. To study authentic integrase in the absence of viral infection in human cells, an ubiquitin fusion strategy was employed to produce any desired residue at the N terminus of the IN protein. (Mulder and Muesing, 2000). The coding sequence of human ubiquitin (Ub) was fused with that of authentic IN to generate Ub-Phe-IN, specifying Phe as the N-terminal amino acid. Integrase fused to Ub were co-translationally cleaved at the Ub-IN junction to produce the Phe-initiated IN. A 3xFLAG-tag was inserted at IN residue 280 (Figure A2.8) to produce tagged Ub-Phe-IN. A corresponding tagged Met initiated IN expression plasmid (Ub-Met-IN-3xFLAG) was also constructed.

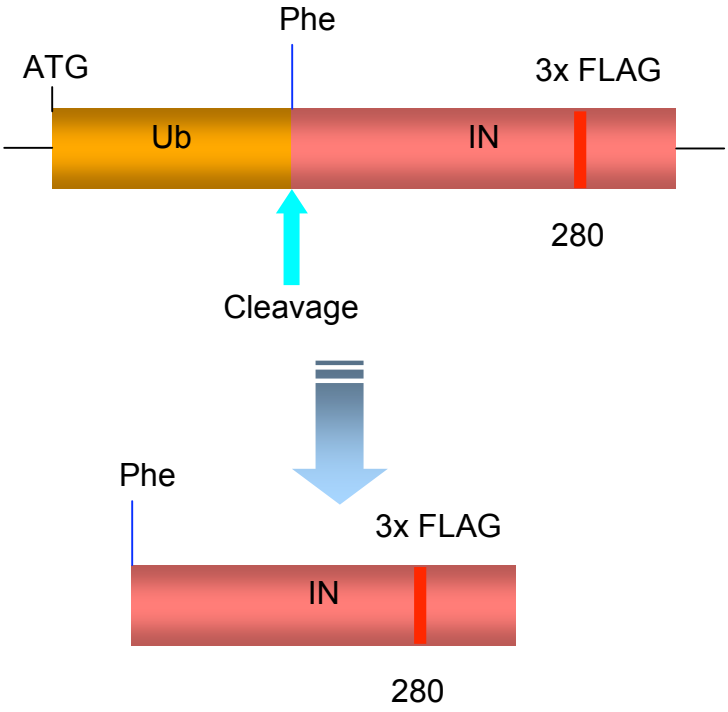


Figure A2.8 Schematic illustration of the HIV-1 integrase expression vector Ub-PheIN-3xFLAG. The arrow denotes the location of the post-translational cleavage between ubiquitin (Ub) and phenylalanine (Phe)-integrase. Ub-Met-IN-3xFLAG was similarly constructed.

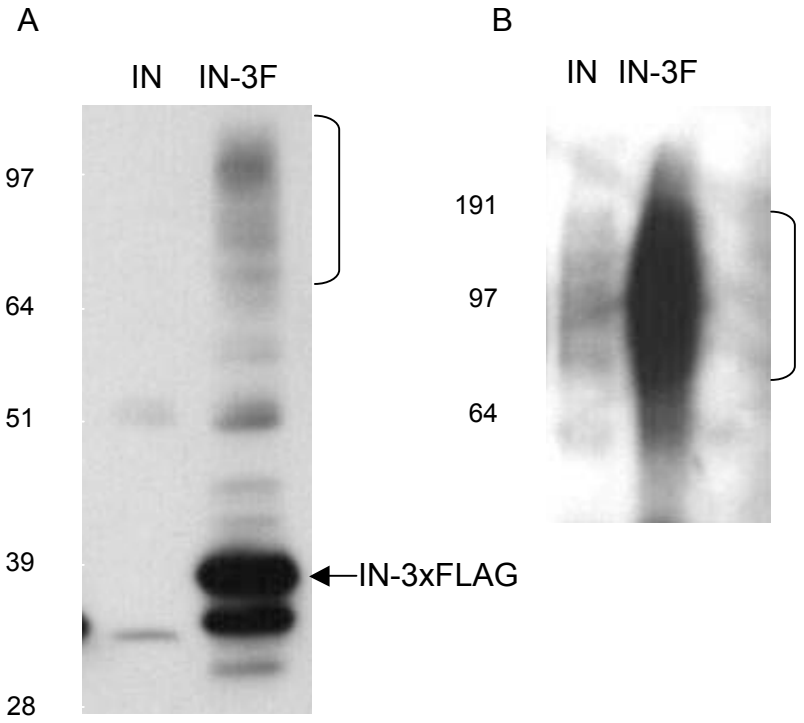


Figure A2.9 Immunoisolation and detection of ubiquitination of the 3xFLAG-tagged (Phe)IN protein in 293T cells. The immunoblot were probed with either (**A**) α-FLAG or (**B**) α-Ub. The smears indicate high molecular weight ubiquitinated products and are bracketed.

Since integrase is degraded by the N-end rule pathway and it interacts with Rad18 in cells, is likely to be ubiquitinated. To test this, Ub-PheIN-3xFLAG was expressed in mammalian cells and immunoprecipitated with α-FLAG. When probed with α-FLAG IN-3xFLAG related products were observed (Figure A2.9A). When probed with α-Ub, the lane with Ub-PheIN-3xFLAG gave a smear at high molecular weight range between 65kDa to 190 kDa, indicating the presence of ubiquitinated (Phe)IN-3xFLAG (Figure A2.9B). We have sought to determine which lysine residue(s) of Ub is preferentially utilized for ubiquitination and the sites within IN targeted for ubiquitin addition. A set of mutant Ub expression vectors were obtained in which all but one Lys residue (at Ub positions: 6, 11, 27, 29, 33, 48 or 63) have been. substituted with Arg (Nishikawa et al., 2004). Each construct was co-transfected with Ub-Phe-IN-3xFLAG and immunoisolated using α-FLAG. Surprisingly, co-transfection of IN-3xFLAG with HA-Ub^{29K} resulted in a band that was about 10 kDa larger than intact IN-3xFLAG (36 kDa) and considerably more intense than any other Ub linkage (Figure A2.10A). When probed with α-HA for the presence of single Lys Ub linked products, the lane with HA-Ub^{29K} gave a different pattern than others. IN-3xFLAG was mainly mono-ubiquitinated rather than poly-ubiquitinated when expressed with Ub^{29K} (Figure A2.10B). There was significant preference for Ub^{29K} to form a mono-ubiquitinated product. In contrast, significantly less mono-ubiquitinated IN and increased level of poly-Ub IN were detected when Phe-IN was expressed as single Lys-to-Arg Ub mutant at position 29 (Ub^{29R}) compared with all other single Lys to Arg mutants (Figure A2.11, lane 4). In addition, this experiment reveals that when the

Ub K48R expression construct is used, a mono-ubiquitinated IN-derived product and reduced the poly-Ub of IN is observed (Figure A2.11, lane 5). Therefore, Ub linkage through residues 48K may result in poly-ubiquitination while Lys-29K preferentially produces a mono-ubiquitinated integrase species.

In attempts to identify ubiquitination sites in integrase, (Phe)IN-3xFLAG was co-transfected with HA-Ub^{29K} since this combination gave the most abundant ubiquitinated product(s). IN-3xFLAG was immunoprecipitated by virtue of the 3xFLAG tag within the IN CTD. The band on SDS-PAGE gel (Figure A2.12) at a molecular weight of 46 kDa was excised and then subjected to MS and MS/MS analysis. Ubiquitin and IN-3xFLAG peptides were identified in the 46 kDa band with by MS. The masses for possible branched Ub fragment, consisting of the acceptor Ub tryptic fragment linked to amino acid GG (or LRGG) from the C-terminus of the donor Ub as well as IN peptide sequences with possible linkage with GG (or LRGG) were calculated and experiment data were searched for matches. However, the site(s) in integrase where ubiquitin is linked to IN was unfortunately inconclusive by MS/MS analysis.

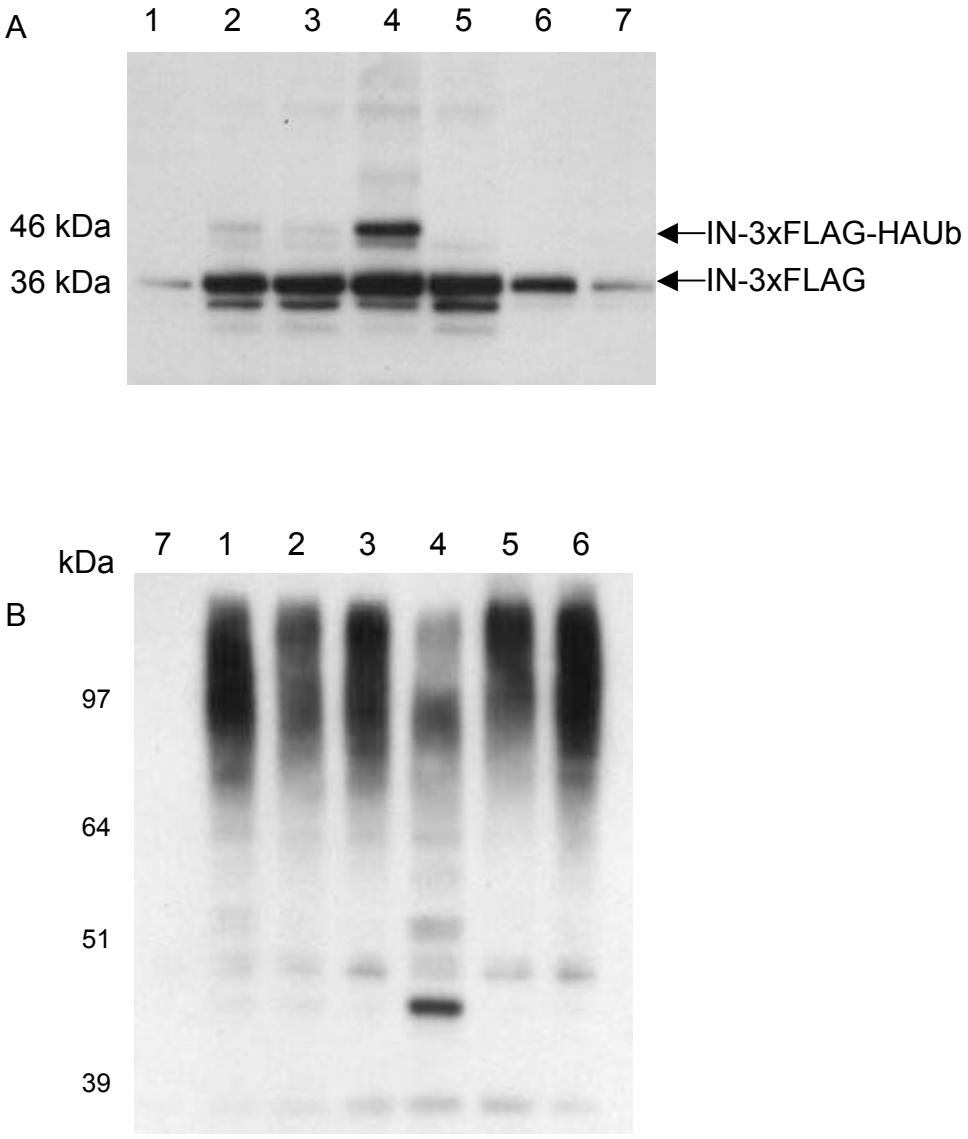


Figure A2.10 Phe-IN-3xFLAG coexpressed with a mutant Ub specifying only a single Lys residue. Ub-Phe-IN-3xFLAG was transfected with the respective HA-tagged Ub mutant expression vector and immunoaffinity purified using α-FLAG. 1: HA-Ub WT; 2: HA-Ub^{6K}; 3: HA-Ub^{11K}; 4: HA-Ub^{29K}; 5: HA-Ub^{48K}; 6: HA-Ub^{63K}; 7: pcDNA3.1. The Western blot was probed with either (**A**) α-FLAG or (**B**) α-HA.

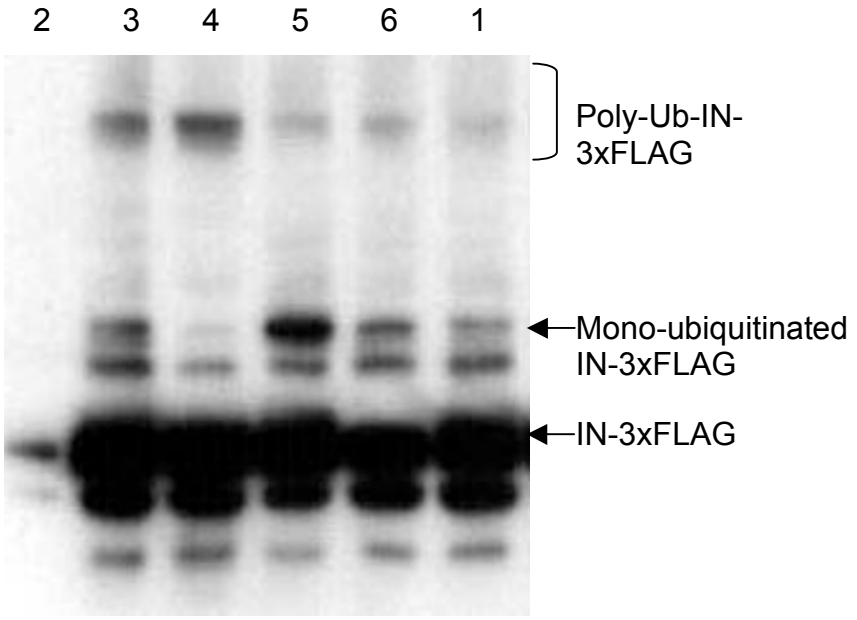


Figure A2.11 Phe-IN-3xFLAG expression with Ub single Lys to Arg substitution mutants. Ub-Phe-IN-3xFLAG was transfected with various HA-Ub mutants and immunoaffinity purified with α-FLAG. 1: Ub wt; 2: Ub^{6R}; 3: Ub^{11R}; 4: Ub^{29R}; 5: Ub^{48R}; 6: Ub^{63R}. The Western blot was probed with α-FLAG.

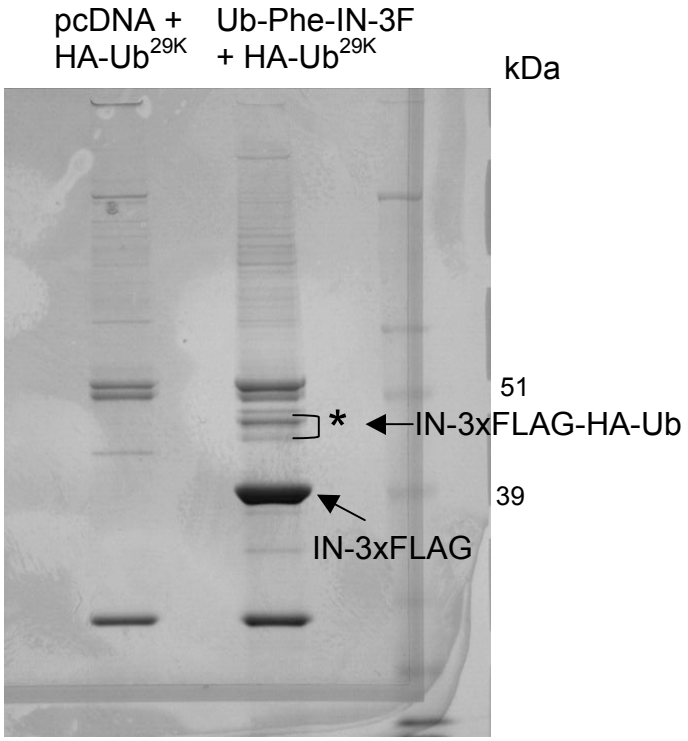


Figure A2.12 Immunoaffinity isolation of ubiquitinated protein complexes using α-FLAG. Ub-Phe-IN-3xFLAG was co-transfected with the HA-Ub^{29K} expression vector. The band representing the mono-ubiquitinated (Phe)IN-3xFLAG species (indicated with *) was excised and processed for MS and MS/MS.

A 2.3 Discussion

A 2.3.1 HAT Acetylation

It is clear from the results that a replication defect occurs only when the 3xFLAG epitope tag is present in conjugation with the K(264/266/273)R (or 3R) mutation, that in primary amino acid sequence, is nearby. The 3R mutation itself does not preclude sustained viral replication in culture. When compared to the WT virus, a genetic assay shows that the integration frequency of the untagged 3R mutant was sufficient to allow replication in culture. Experiments by others in the lab confirm that the IN CTD is an efficient target for HAT-mediated acetylation by p300 *in vitro* (Topper et al., 2006). In addition, the steady-state level of late RT products synthesized during 3R mutant infection compared to WT closely paralleled the results we obtained using other assays designed to detect any observable defects associated with the 3R mutation. Taken together, these determinations cast doubt on the significance of acetylation (or other post-translational modification) at these specific lysine residues (264, 266, 273) in the immortalized cell types examined. However, the three lysines targeted for acetylation are conserved across all clades of HIV-1 and therefore suggest that their preservation is important for some aspect of viral replication. For example, the specific lysine at position 273 might be important for maintaining the integrity of the Vif coding frame since the 273 codon is within the *pol-vif* overlap region (Topper et al., 2006). Despite this, all three residues can be conservatively replaced with arginine in single, double or triple combinations while remaining proficient for IN function and viral replication. This suggests that acetylation is not

essential for IN function during infection of immortalized cells. Thus, we cannot confirm the original observation of the Cereseto group that post-translational acetylation by p300 is essential for viral integration and replication. The discrepancy between our work and that of the Cereseto and coworkers may be explained by their use of an acidic epitope tag placed at the extreme carboxyl terminus of integrase, near the target site for acetylation and mutation (Topper et al., 2006). Indeed, the combination of this epitope tag with the 3R substitution results in a replication-defective phenotype while the untagged 3R mutant is a replication competent virus. These observations provide evidence for an unanticipated experimental caveat and highlight the need for caution when interpreting data generated from tagged viruses and their mutant derivatives.

A2.3.2 Casein Kinase II

CKII is shown here to phosphorylate HIV-1 integrase *in vitro*. We have characterized the domain and phosphorylation sites responsible for reactivity with the CKII kinase. Study using IN truncations mapped the IN-CTD (212-288) to be the major participating substrate for CKII phosphorylation. The serine at residue 283 is the major target of CKII-mediated phosphorylation *in vitro*. Substitution of this serine (S283A) diminished, but did not abolish, CKII phosphorylation. This indicates that other sites in IN are utilized for CKII phosphorylation, although perhaps through promiscuous activity of this kinase. In attempts to map all the CKII sites in IN, mass spectrometry confirmed that Ser 283 is indeed phosphorylated. However, MS/MS analysis of other potential phosphorylation

sites was inconclusive due to poor recovery of the IN peptide segment at residues 50-120.

Although S283 reacts with CKII *in vitro*, the biological relevance of this observation remains to be clarified. IN S283A does not show significant change in subcellular localization when compared to nuclear localized WT IN. Ser to Ala single mutation at IN residue 283 had no major effect on viral replication. It is possible that multiple sites within IN participate in phosphorylation events, and the effect of a single mutation is masked by the function of others. Furthermore, we are unable to demonstrate any interaction between IN and CKII *in vivo* using immunoprecipitation protocols under any of the conditions examined. Although physical association might not be necessary for CKII phosphorylation, optimization of the lysis conditions could be critical.

A2.3.3 Ubiquitination

HIV-1 IN is the substrate of N-end rule pathway and it interacts with Rad18, an E3 Ub ligase (Mulder et al., 2002; Mulder and Muesing, 2000; Tasaki et al., 2005). However, the temporal and specific requirement of IN ubiquitination in the context of viral infection has yet to be characterized. The studies reported here examine IN ubiquitination when IN is expressed in cells in the absence of other viral proteins. In these experiments, IN is ubiquitinated in mammalian cells. We identify covalent lysine linkage of both mono- and poly-Ub species to IN. Our results show that the pattern of IN ubiquitination changes dramatically depending

upon the use of Ub-Lys-29K or Ub-Lys-48R. IN preferentially utilizes Ub^{29K} to produce a mono-ubiquitinated product *in vivo* (Figure A2.10), and mutation at Ub-K29R severely reduces the level of a mono-Ub-IN product (Figure A2.11). This may indicate that mono-Ub of IN is catalyzed using the Ub-Lys-29 residue. Another possibility is that other Lys linkages, especially 48K, results in the poly-Ub of IN. Indeed, K48R mutation in Ub reduces poly-Ub products with concomitant increases in the mono-Ub form of IN. When a Ub mutant in which the only available Lys is at position 48 is used, it yields predominantly poly-ubiquitinated products, indicating that IN may preferentially utilize Ub^{48K} for poly-ubiquitination. The significance of possible mono-Ub formation using Ub^{29K} is not clear at present, although mono-ubiquitination has been shown to prevent intramolecular interactions of the modified protein with other ubiquitinated targets (Hoeller et al., 2006). On the other hand, the most common poly-Ub chain is linked through Lys-48 of Ub and serves as a signal for rapid degradation of substrates (Chau et al., 1989).

For many proteins, mass spectrometry analysis has been of great utility to identify Ub linkage(s) required for mono- or poly-ubiquitination. Here, using MS, monomeric Ub was shown to be covalently linked to IN. However, the results of MS/MS for details regarding the preferred and precise sites within IN targeted for Ub modification were inconclusive because of the low level of the isolated Ub-IN. It would be invaluable if we can recovery integrase peptide(s) that linked with Ub (addition of G-G after trypsin digestion) or Lys-48-linked Ub fragment from poly-

Ub IN.

In summary, our studies on integrase post-translational modifications (PTMs) including acetylation, phosphorylation and ubiquitination have identified potential PTM sites. For some of the presumptive sites targeted for modification, the behavior of the relevant mutants of the IN protein expressed in isolation or within the context of viral replication have been characterized. Further experiments are warranted to decipher the possible implications of IN modification to HIV-1 replication in the more clinically relevant cell types.

References

Alvisi, G., Jans, D.A., Guo, J., Pinna, L.A., and Ripalti, A. (2005). A protein kinase CK2 site flanking the nuclear targeting signal enhances nuclear transport of human cytomegalovirus ppUL44. Traffic (Copenhagen, Denmark) *6*, 1002-1013.

Ao, Z., Huang, G., Yao, H., Xu, Z., Labine, M., Cochrane, A.W., and Yao, X. (2007). Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. The Journal of biological chemistry.

Argasinska, J., Zhou, K., Donnelly, R.J., Hay, R.T., and Lee, C.G. (2004). A functional interaction between RHA and Ubc9, an E2-like enzyme specific for Sumo-1. Journal of molecular biology *341*, 15-25.

Arhel, N., Genovesio, A., Kim, K.A., Miko, S., Perret, E., Olivo-Marin, J.C., Shorte, S., and Charneau, P. (2006). Quantitative four-dimensional tracking of cytoplasmic and nuclear HIV-1 complexes. Nature methods *3*, 817-824.

Ariumi, Y., Turelli, P., Masutani, M., and Trono, D. (2005). DNA damage sensors ATM, ATR, DNA-PKcs, and PARP-1 are dispensable for human immunodeficiency virus type 1 integration. J Virol *79*, 2973-2978.

Armon-Omer, A., Graessmann, A., and Loyter, A. (2004). A synthetic peptide bearing the HIV-1 integrase 161-173 amino acid residues mediates active nuclear import and binding to importin alpha: characterization of a functional nuclear localization signal. Journal of molecular biology *336*, 1117-1128.

Asante-Appiah, E., and Skalka, A.M. (1997). Molecular mechanisms in retrovirus DNA integration. Antiviral research *36*, 139-156.

Atasheva, S., Gorchakov, R., English, R., Frolov, I., and Frolova, E. (2007). Development of Sindbis viruses encoding nsP2/GFP chimeric protein and their application for studying nsP2 functioning. J Virol.

Baekelandt, V., Claeys, A., Cherepanov, P., De Clercq, E., De Strooper, B., Nuttin, B., and Debyser, Z. (2000). DNA-Dependent protein kinase is not required for efficient lentivirus integration. J Virol *74*, 11278-11285.

Bailly, V., Lauder, S., Prakash, S., and Prakash, L. (1997). Yeast DNA repair proteins Rad6 and Rad18 form a heterodimer that has ubiquitin conjugating, DNA binding, and ATP hydrolytic activities. The Journal of biological chemistry *272*, 23360-23365.

Barany, F. (1985). Two-codon insertion mutagenesis of plasmid genes by using single-stranded hexameric oligonucleotides. Proceedings of the National Academy of Sciences of the United States of America *82*, 4202-4206.

Baumann, H., Knapp, S., Lundback, T., Ladenstein, R., and Hard, T. (1994). Solution structure and DNA-binding properties of a thermostable protein from the archaeon Sulfolobus solfataricus. Nat Struct Biol *1*, 808-819.

Biery, M.C., Stewart, F.J., Stellwagen, A.E., Raleigh, E.A., and Craig, N.L. (2000). A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. Nucleic acids research *28*, 1067-1077.

Bizub-Bender, D., Kulkosky, J., and Skalka, A.M. (1994). Monoclonal antibodies against HIV type 1 integrase: clues to molecular structure. AIDS research and human retroviruses *10*, 1105-1115.

Bonifacino, J.S., and Weissman, A.M. (1998). Ubiquitin and the control of protein fate in the secretory and endocytic pathways. Annual review of cell and developmental biology *14*, 19-57.

Bouyac-Bertoia, M., Dvorin, J.D., Fouchier, R.A., Jenkins, Y., Meyer, B.E., Wu, L.I., Emerman, M., and Malim, M.H. (2001). HIV-1 infection requires a functional integrase NLS. Molecular cell *7*, 1025-1035.

Brandt, S.M., Mariani, R., Holland, A.U., Hope, T.J., and Landau, N.R. (2002). Association of chemokine-mediated block to HIV entry with coreceptor internalization. The Journal of biological chemistry *277*, 17291-17299.

Bruce, J.W., Bradley, K.A., Ahlquist, P., and Young, J.A. (2005). Isolation of cell lines that show novel, murine leukemia virus-specific blocks to early steps of retroviral replication. J Virol *79*, 12969-12978.

Brune, W., Menard, C., Hobom, U., Odenbreit, S., Messerle, M., and Koszinowski, U.H. (1999). Rapid identification of essential and nonessential herpesvirus genes by direct transposon mutagenesis. Nature biotechnology *17*, 360-364.

Buranda, T., Lopez, G.P., Simons, P., Pastuszyn, A., and Sklar, L.A. (2001). Detection of epitope-tagged proteins in flow cytometry: fluorescence resonance energy transfer-based assays on beads with femtomole resolution. Analytical biochemistry *298*, 151-162.

Burton, D.R., Desrosiers, R.C., Doms, R.W., Koff, W.C., Kwong, P.D., Moore, J.P., Nabel, G.J., Sodroski, J., Wilson, I.A., and Wyatt, R.T. (2004). HIV vaccine design and the neutralizing antibody problem. Nature immunology *5*, 233-236.

Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ, F., and Debyser, Z. (2005). The interaction of LEDGF/p75 with integrase

is lentivirus-specific and promotes DNA binding. The Journal of biological chemistry *280*, 17841-17847.

Cao, Y., Qin, L., Zhang, L., Safrit, J., and Ho, D.D. (1995). Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection. The New England journal of medicine *332*, 201-208.

Cavacini, L., and Posner, M. (2004). Native HIV type 1 virion surface structures: relationships between antibody binding and neutralization or lessons from the viral capture assay. AIDS research and human retroviruses *20*, 435-441.

Cereseto, A., Manganaro, L., Gutierrez, M.I., Terreni, M., Fittipaldi, A., Lusic, M., Marcello, A., and Giacca, M. (2005). Acetylation of HIV-1 integrase by p300 regulates viral integration. The EMBO journal *24*, 3070-3081.

Charneau, P., Mirambeau, G., Roux, P., Paulous, S., Buc, H., and Clavel, F. (1994). HIV-1 reverse transcription. A termination step at the center of the genome. Journal of molecular biology *241*, 651-662.

Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K., and Varshavsky, A. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science *243*, 1576-1583.

Chen, J.C., Krucinski, J., Miercke, L.J., Finer-Moore, J.S., Tang, A.H., Leavitt, A.D., and Stroud, R.M. (2000). Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. Proceedings of the National Academy of Sciences of the United States of America *97*, 8233-8238.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. The Journal of biological chemistry *278*, 372-381.

Cherepanov, P., Pluymers, W., Claeys, A., Proost, P., De Clercq, E., and Debyser, Z. (2000). High-level expression of active HIV-1 integrase from a synthetic gene in human cells. Faseb J *14*, 1389-1399.

Chilkoti, A., Tan, P.H., and Stayton, P.S. (1995). Site-directed mutagenesis studies of the high-affinity streptavidin-biotin complex: contributions of tryptophan residues 79, 108, and 120. Proceedings of the National Academy of Sciences of the United States of America *92*, 1754-1758.

Coffin, J.M. (1992). Genetic diversity and evolution of retroviruses. Current topics in microbiology and immunology *176*, 143-164.

Coffin, J.M., Hughes, S.H., and Varmus, H. (1997). Retroviruses (Plainview, N.Y., Cold Spring Harbor Laboratory Press).

Cohen, P. (2000). The regulation of protein function by multisite phosphorylation--a 25 year update. Trends in biochemical sciences *25*, 596-601.

Conley, A.J., Gorny, M.K., Kessler, J.A., 2nd, Boots, L.J., Ossorio-Castro, M., Koenig, S., Lineberger, D.W., Emini, E.A., Williams, C., and Zolla-Pazner, S. (1994). Neutralization of primary human immunodeficiency virus type 1 isolates by the broadly reactive anti-V3 monoclonal antibody, 447-52D. J Virol *68*, 6994-7000.

Craig, N.L. (1996). Transposon Tn7. Current topics in microbiology and immunology *204*, 27-48.

Cristea, I.M., Carroll, J.W., Rout, M.P., Rice, C.M., Chait, B.T., and MacDonald, M.R. (2006). Tracking and elucidating alphavirus-host protein interactions. The Journal of biological chemistry *281*, 30269-30278.

Cristea, I.M., Williams, R., Chait, B.T., and Rout, M.P. (2005). Fluorescent proteins as proteomic probes. Mol Cell Proteomics *4*, 1933-1941.

Daniel, R., Katz, R.A., and Skalka, A.M. (1999). A role for DNA-PK in retroviral DNA integration. Science *284*, 644-647.

Darlix, J.L., Lapadat-Tapolsky, M., de Rocquigny, H., and Roques, B.P. (1995). First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. Journal of molecular biology *254*, 523-537.

Deiss, L.P., and Kimchi, A. (1991). A genetic tool used to identify thioredoxin as a mediator of a growth inhibitory signal. Science *252*, 117-120.

di Guan, C., Li, P., Riggs, P.D., and Inouye, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. Gene *67*, 21-30.

Ehrlich, E.S., and Yu, X.F. (2006). Lentiviral Vif: viral hijacker of the ubiquitin-proteasome system. International journal of hematology *83*, 208-212.

Eijkelenboom, A.P., Lutzke, R.A., Boelens, R., Plasterk, R.H., Kaptein, R., and Hard, K. (1995). The DNA-binding domain of HIV-1 integrase has an SH3-like fold. Nat Struct Biol *2*, 807-810.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.

Emiliani, S., Mousnier, A., Busschots, K., Maroun, M., Van Maele, B., Tempe, D., Vandekerckhove, L., Moisant, F., Ben-Slama, L., Witvrouw, M.*, et al.* (2005). Integrase mutants defective for interaction with LEDGF/p75 are impaired in

chromosome tethering and HIV-1 replication. The Journal of biological chemistry *280*, 25517-25523.

Engelman, A., Hickman, A.B., and Craigie, R. (1994). The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. J Virol *68*, 5911-5917.

Evans, M.J., von Hahn, T., Tscherne, D.M., Syder, A.J., Panis, M., Wolk, B., Hatziioannou, T., McKeating, J.A., Bieniasz, P.D., and Rice, C.M. (2007). Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. Nature.

Fackler, O.T., and Krausslich, H.G. (2006). Interactions of human retroviruses with the host cell cytoskeleton. Curr Opin Microbiol *9*, 409-415.

Fraser, A.G., Kamath, R.S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000). Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. Nature *408*, 325-330.

Freed, E.O. (1998). HIV-1 gag proteins: diverse functions in the virus life cycle. Virology *251*, 1-15.

Friborg, J., Ladha, A., Gottlinger, H., Haseltine, W.A., and Cohen, E.A. (1995). Functional analysis of the phosphorylation sites on the human immunodeficiency virus type 1 Vpu protein. J Acquir Immune Defic Syndr Hum Retrovirol *8*, 10-22.

Fujii, R., Okamoto, M., Aratani, S., Oishi, T., Ohshima, T., Taira, K., Baba, M., Fukamizu, A., and Nakajima, T. (2001). A Role of RNA Helicase A in cis-Acting Transactivation Response Element-mediated Transcriptional Regulation of Human Immunodeficiency Virus Type 1. The Journal of biological chemistry *276*, 5445-5451.

Fung, M.S., Sun, C.R., Gordon, W.L., Liou, R.S., Chang, T.W., Sun, W.N., Daar, E.S., and Ho, D.D. (1992). Identification and characterization of a neutralization site within the second variable region of human immunodeficiency virus type 1 gp120. J Virol *66*, 848-856.

Gallay, P., Hope, T., Chin, D., and Trono, D. (1997). HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. Proceedings of the National Academy of Sciences of the United States of America *94*, 9825-9830.

Gao, G., and Goff, S.P. (1999). Somatic cell mutants resistant to retrovirus replication: intracellular blocks during the early stages of infection. Molecular biology of the cell *10*, 1705-1717.

Gao, K., Butler, S.L., and Bushman, F. (2001). Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes. The EMBO journal *20*, 3565-3576.

Garrus, J.E., von Schwedler, U.K., Pornillos, O.W., Morham, S.G., Zavitz, K.H., Wang, H.E., Wettstein, D.A., Stray, K.M., Cote, M., Rich, R.L.*, et al.* (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell *107*, 55-65.

Gatfield, D., Le Hir, H., Schmitt, C., Braun, I.C., Kocher, T., Wilm, M., and Izaurralde, E. (2001). The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. Curr Biol *11*, 1716-1721.

Gatignol, A., Buckler-White, A., Berkhout, B., and Jeang, K.T. (1991). Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. Science *251*, 1597-1600.

Gesellchen, F., Bertinetti, O., and Herberg, F.W. (2006). Analysis of posttranslational modifications exemplified using protein kinase A. Biochim Biophys Acta *1764*, 1788-1800.

Gianni, A.M., Smotkin, D., and Weinberg, R.A. (1975). Murine leukemia virus: detection of unintegrated double-stranded DNA forms of the provirus. Proceedings of the National Academy of Sciences of the United States of America *72*, 447-451.

Glover, C.V., 3rd (1998). On the physiological role of casein kinase II in Saccharomyces cerevisiae. Progress in nucleic acid research and molecular biology *59*, 95-133.

Goff, S.P. (2007). Host factors exploited by retroviruses. Nat Rev Microbiol *5*, 253-263.

Goff, S.P., and Prasad, V.R. (1991). Linker insertion mutagenesis as probe of structure-function relationships. Methods in enzymology *208*, 586-603.

Goncalves, J., Jallepalli, P., and Gabuzda, D.H. (1994). Subcellular localization of the Vif protein of human immunodeficiency virus type 1. J Virol *68*, 704-712.

Gorny, M.K., Conley, A.J., Karwowska, S., Buchbinder, A., Xu, J.Y., Emini, E.A., Koenig, S., and Zolla-Pazner, S. (1992). Neutralization of diverse human immunodeficiency virus type 1 variants by an anti-V3 human monoclonal antibody. J Virol *66*, 7538-7542.

Greenspan, N.S. (2001). Affinity, complementarity, cooperativity, and specificity in antibody recognition. Current topics in microbiology and immunology *260*, 65-85.

Gu, W., and Roeder, R.G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell *90*, 595-606.

Hamamoto, S., Nishitsuji, H., Amagasa, T., Kannagi, M., and Masuda, T. (2006). Identification of a novel human immunodeficiency virus type 1 integrase interactor, Gemin2, that facilitates efficient viral cDNA synthesis in vivo. J Virol *80*, 5670-5677.

Hannon, G.J., Sun, P., Carnero, A., Xie, L.Y., Maestro, R., Conklin, D.S., and Beach, D. (1999). MaRX: an approach to genetics in mammalian cells. Science *283*, 1129-1130.

Hearps, A.C., and Jans, D.A. (2006). HIV-1 integrase is capable of targeting DNA to the nucleus via an importin alpha/beta-dependent mechanism. The Biochemical journal *398*, 475-484.

Hernan, R., Heuermann, K., and Brizzard, B. (2000). 3x FLAG: Ultra-Sensitive Detection of Recombinant Proteins. A Technical Application Newsletter *1*.

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. Annual review of biochemistry *67*, 425-479.

Hicke, L. (2001). Protein regulation by monoubiquitin. Nature reviews *2*, 195-201.

Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. Journal of molecular biology *26*, 365-369.

Hoege, C., Pfander, B., Moldovan, G.L., Pyrowolakis, G., and Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature *419*, 135-141.

Hoeller, D., Crosetto, N., Blagoev, B., Raiborg, C., Tikkanen, R., Wagner, S., Kowanetz, K., Breitling, R., Mann, M., Stenmark, H.*, et al.* (2006). Regulation of ubiquitin-binding proteins by monoubiquitination. Nature cell biology *8*, 163-169.

Hopp, T.P., Gallis, B., and Prickett, K.S. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. Bio/Technology *6*, 1204–1210.

Horton, R., Mumm, S.R., and Grandgenett, D.P. (1991). Phosphorylation of the avian retrovirus integration protein and proteolytic processing of its carboxyl terminus. J Virol *65*, 1141-1148.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nature biotechnology *21*, 635-637.

Jares-Erijman, E.A., and Jovin, T.M. (2006). Imaging molecular interactions in living cells by FRET microscopy. Curr Opin Chem Biol *10*, 409-416.

Jeffery, C.J. (2003). Moonlighting proteins: old proteins learning new tricks. Trends Genet *19*, 415-417.

Jenkins, T.M., Hickman, A.B., Dyda, F., Ghirlando, R., Davies, D.R., and Craigie, R. (1995). Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. Proceedings of the National Academy of Sciences of the United States of America *92*, 6057-6061.

Joyner, A.L., Skarnes, W.C., and Rossant, J. (1989). Production of a mutation in mouse En-2 gene by homologous recombination in embryonic stem cells. Nature *338*, 153-156.

Jurriaans, S., de Ronde, A., Dekker, J., Goudsmit, J., and Cornelissen, M. (1992). Analysis of human immunodeficiency virus type 1 LTR-LTR junctions in peripheral blood mononuclear cells of infected individuals. The Journal of general virology *73 (Pt 6)*, 1537-1541.

Kahn, R.A., Fu, H., and Roy, C.R. (2002). Cellular hijacking: a common strategy for microbial infection. Trends in biochemical sciences *27*, 308-314.

Kalpana, G.V., Marmon, S., Wang, W., Crabtree, G.R., and Goff, S.P. (1994). Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science *266*, 2002-2006.

Kannouche, P.L., Wing, J., and Lehmann, A.R. (2004). Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. Molecular cell *14*, 491-500.

Karczewski, M.K., and Strebel, K. (1996). Cytoskeleton association and virion incorporation of the human immunodeficiency virus type 1 Vif protein. J Virol *70*, 494-507.

Kikkawa, U., Mann, S.K., Firtel, R.A., and Hunter, T. (1992). Molecular cloning of casein kinase II alpha subunit from Dictyostelium discoideum and its expression in the life cycle. Molecular and cellular biology *12*, 5711-5723.

Kolchinsky, P., Kiprilov, E., Bartley, P., Rubinstein, R., and Sodroski, J. (2001). Loss of a single N-linked glycan allows CD4-independent human immunodeficiency virus type 1 infection by altering the position of the gp120 V1/V2 variable loops. J Virol *75*, 3435-3443.

Krutchinsky, A.N., Kalkum, M., and Chait, B.T. (2001). Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. Analytical chemistry *73*, 5066-5077.

Kuiken, C., Korber, B., and Shafer, R.W. (2003). HIV sequence databases. AIDS reviews *5*, 52-61.

Kumar, A., Seringhaus, M., Biery, M.C., Sarnovsky, R.J., Umansky, L., Piccirillo, S., Heidtman, M., Cheung, K.H., Dobry, C.J., Gerstein, M.B.*, et al.* (2004). Large-scale mutagenesis of the yeast genome using a Tn7-derived multipurpose transposon. Genome research *14*, 1975-1986.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature *393*, 648-659.

Labrijn, A.F., Poignard, P., Raja, A., Zwick, M.B., Delgado, K., Franti, M., Binley, J., Vivona, V., Grundner, C., Huang, C.C.*, et al.* (2003). Access of antibody molecules to the conserved coreceptor binding site on glycoprotein gp120 is sterically restricted on primary human immunodeficiency virus type 1. J Virol *77*, 10557-10565.

LaFemina, R.L., Schneider, C.L., Robbins, H.L., Callahan, P.L., LeGrow, K., Roth, E., Schleif, W.A., and Emini, E.A. (1992). Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells. J Virol *66*, 7414-7419.

Landesman-Bollag, E., Romieu-Mourez, R., Song, D.H., Sonenshein, G.E., Cardiff, R.D., and Seldin, D.C. (2001). Protein kinase CK2 in mammary gland tumorigenesis. Oncogene *20*, 3247-3257.

Larsen, M.R., Trelle, M.B., Thingholm, T.E., and Jensen, O.N. (2006). Analysis of posttranslational modifications of proteins by tandem mass spectrometry. BioTechniques *40*, 790-798.

Laurent, L.C., Olsen, M.N., Crowley, R.A., Savilahti, H., and Brown, P.O. (2000). Functional characterization of the human immunodeficiency virus type 1 genome by genetic footprinting. J Virol *74*, 2760-2769.

Lee, C.G., and Hurwitz, J. (1992). A new RNA helicase isolated from HeLa cells that catalytically translocates in the 3' to 5' direction. The Journal of biological chemistry *267*, 4398-4407.

Li, J., Tang, H., Mullen, T.M., Westberg, C., Reddy, T.R., Rose, D.W., and Wong-Staal, F. (1999). A role for RNA helicase A in post-transcriptional regulation of HIV type 1. Proceedings of the National Academy of Sciences of the United States of America *96*, 709-714.

Li, L., Olvera, J.M., Yoder, K.E., Mitchell, R.S., Butler, S.L., Lieber, M., Martin, S.L., and Bushman, F.D. (2001). Role of the non-homologous DNA end joining

pathway in the early steps of retroviral infection. The EMBO journal *20*, 3272-3281.

Li, M., Brooks, C.L., Wu-Baer, F., Chen, D., Baer, R., and Gu, W. (2003). Mono- versus polyubiquitination: differential control of p53 fate by Mdm2. Science *302*, 1972-1975.

Litchfield, D.W. (2003). Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. The Biochemical journal *369*, 1-15.

Llano, M., Saenz, D.T., Meehan, A., Wongthida, P., Peretz, M., Walker, W.H., Teo, W., and Poeschla, E.M. (2006). An essential role for LEDGF/p75 in HIV integration. Science *314*, 461-464.

Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M., and Poeschla, E.M. (2004). 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 *78*, 9524-9537.

Lu, R., Ghory, H.Z., and Engelman, A. (2005). Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase. J Virol *79*, 10356-10368.

Lutzke, R.A., and Plasterk, R.H. (1998). 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 *72*, 4841-4848.

Ma, Y., Creanga, A., Lum, L., and Beachy, P.A. (2006). Prevalence of off-target effects in Drosophila RNA interference screens. Nature *443*, 359-363.

Maertens, G., Cherepanov, P., Debyser, Z., Engelborghs, Y., and Engelman, A. (2004). Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75. The Journal of biological chemistry *279*, 33421-33429.

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z., and Engelborghs, Y. (2003). LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. The Journal of biological chemistry *278*, 33528-33539.

Manes, S., del Real, G., and Martinez, A.C. (2003). Pathogens: raft hijackers. Nat Rev Immunol *3*, 557-568.

Mann, M., and Jensen, O.N. (2003). Proteomic analysis of post-translational modifications. Nature biotechnology *21*, 255-261.

McDonald, D., Vodicka, M.A., Lucero, G., Svitkina, T.M., Borisy, G.G., Emerman, M., and Hope, T.J. (2002). Visualization of the intracellular behavior of HIV in living cells. The Journal of cell biology *159*, 441-452.

Meggio, F., Marin, O., Boschetti, M., Sarno, S., and Pinna, L.A. (2001). HIV-1 Rev transactivator: a beta-subunit directed substrate and effector of protein kinase CK2. Mol Cell Biochem *227*, 145-151.

Miyase, S., Tateishi, S., Watanabe, K., Tomita, K., Suzuki, K., Inoue, H., and Yamaizumi, M. (2005). Differential regulation of Rad18 through Rad6-dependent mono- and polyubiquitination. The Journal of biological chemistry *280*, 515-524.

Moradpour, D., Evans, M.J., Gosert, R., Yuan, Z., Blum, H.E., Goff, S.P., Lindenbach, B.D., and Rice, C.M. (2004). Insertion of green fluorescent protein into nonstructural protein 5A allows direct visualization of functional hepatitis C virus replication complexes. J Virol *78*, 7400-7409.

Morner, A., Bjorndal, A., Albert, J., Kewalramani, V.N., Littman, D.R., Inoue, R., Thorstensson, R., Fenyo, E.M., and Bjorling, E. (1999). Primary human immunodeficiency virus type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in coreceptor usage. J Virol *73*, 2343-2349.

Mulder, L.C., Chakrabarti, L.A., and Muesing, M.A. (2002). Interaction of HIV-1 integrase with DNA repair protein hRad18. The Journal of biological chemistry *277*, 27489-27493.

Mulder, L.C., and Muesing, M.A. (2000). Degradation of HIV-1 integrase by the N-end rule pathway. The Journal of biological chemistry *275*, 29749-29753.

Nakajima, T., Uchida, C., Anderson, S.F., Lee, C.G., Hurwitz, J., Parvin, J.D., and Montminy, M. (1997). RNA helicase A mediates association of CBP with RNA polymerase II. Cell *90*, 1107-1112.

Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., and Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science *272*, 263-267.

Navaratnarajah, C.K., and Kuhn, R.J. (2007). Functional characterization of the Sindbis virus E2 glycoprotein by transposon linker-insertion mutagenesis. Virology.

Neville, M., Stutz, F., Lee, L., Davis, L.I., and Rosbash, M. (1997). The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. Curr Biol *7*, 767-775.

Nilsen, B.M., Haugan, I.R., Berg, K., Olsen, L., Brown, P.O., and Helland, D.E. (1996). Monoclonal antibodies against human immunodeficiency virus type 1

integrase: epitope mapping and differential effects on integrase activities in vitro. J Virol *70*, 1580-1587.

Nishikawa, H., Ooka, S., Sato, K., Arima, K., Okamoto, J., Klevit, R.E., Fukuda, M., and Ohta, T. (2004). Mass spectrometric and mutational analyses reveal Lys-6-linked polyubiquitin chains catalyzed by BRCA1-BARD1 ubiquitin ligase. The Journal of biological chemistry *279*, 3916-3924.

Parren, P.W., Mondor, I., Naniche, D., Ditzel, H.J., Klasse, P.J., Burton, D.R., and Sattentau, Q.J. (1998). Neutralization of human immunodeficiency virus type 1 by antibody to gp120 is determined primarily by occupancy of sites on the virion irrespective of epitope specificity. J Virol *72*, 3512-3519.

Petit, C., Schwartz, O., and Mammano, F. (1999). Oligomerization within virions and subcellular localization of human immunodeficiency virus type 1 integrase. J Virol *73*, 5079-5088.

Pfeiffer, P., and Schum, D. (1973). Introduction to Applied Probability (New York, Academic Press).

Pickart, C.M. (2001). Mechanisms underlying ubiquitination. Annual review of biochemistry *70*, 503-533.

Potts, B. (1990). "Mini" reverse transcriptase assay (New York, NY, Stockton Press).

Prickett, K.S., Amberg, D.C., and Hopp, T.P. (1989). A calcium-dependent antibody for identification and purification of recombinant proteins. BioTechniques *7*, 580-589.

Quinones-Mateu, M.E., Ball, S.C., Marozsan, A.J., Torre, V.S., Albright, J.L., Vanham, G., van Der Groen, G., Colebunders, R.L., and Arts, E.J. (2000). A dual infection/competition assay shows a correlation between ex vivo human immunodeficiency virus type 1 fitness and disease progression. J Virol *74*, 9222-9233.

Ragolia, L., Hall, C.E., and Palaia, T. (2007). Post-translational modification regulates prostaglandin D2 synthase apoptotic activity: characterization by site-directed mutagenesis. Prostaglandins Other Lipid Mediat *83*, 25-32.

Reiser, J., Harmison, G., Kluepfel-Stahl, S., Brady, R.O., Karlsson, S., and Schubert, M. (1996). Transduction of nondividing cells using pseudotyped defective high-titer HIV type 1 particles. Proceedings of the National Academy of Sciences of the United States of America *93*, 15266-15271.

Ren, X., Sodroski, J., and Yang, X. (2005). An unrelated monoclonal antibody neutralizes human immunodeficiency virus type 1 by binding to an artificial

epitope engineered in a functionally neutral region of the viral envelope glycoproteins. J Virol *79*, 5616-5624.

Resh, M.D. (2006). Trafficking and signaling by fatty-acylated and prenylated proteins. Nat Chem Biol *2*, 584-590.

Rits-Volloch, S., Frey, G., Harrison, S.C., and Chen, B. (2006). Restraining the conformation of HIV-1 gp120 by removing a flexible loop. The EMBO journal *25*, 5026-5035.

Rocak, S., and Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. Nature reviews *5*, 232-241.

Roos, J.W., Maughan, M.F., Liao, Z., Hildreth, J.E., and Clements, J.E. (2000). LuSIV cells: a reporter cell line for the detection and quantitation of a single cycle of HIV and SIV replication. Virology *273*, 307-315.

Roth, M.J., Schwartzberg, P., Tanese, N., and Goff, S.P. (1990). Analysis of mutations in the integration function of Moloney murine leukemia virus: effects on DNA binding and cutting. J Virol *64*, 4709-4717.

Rothenberg, S.M., Olsen, M.N., Laurent, L.C., Crowley, R.A., and Brown, P.O. (2001). Comprehensive mutational analysis of the Moloney murine leukemia virus envelope protein. J Virol *75*, 11851-11862.

Roy, B.B., Hu, J., Guo, X., Russell, R.S., Guo, F., Kleiman, L., and Liang, C. (2006). Association of RNA helicase a with human immunodeficiency virus type 1 particles. The Journal of biological chemistry *281*, 12625-12635.

Scacheri, P.C., Rozenblatt-Rosen, O., Caplen, N.J., Wolfsberg, T.G., Umayam, L., Lee, J.C., Hughes, C.M., Shanmugam, K.S., Bhattacharjee, A., Meyerson, M.*, et al.* (2004). Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *101*, 1892-1897.

Schrofelbauer, B., Chen, D., and Landau, N.R. (2004). A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proceedings of the National Academy of Sciences of the United States of America *101*, 3927-3932.

Schrofelbauer, B., Hakata, Y., and Landau, N.R. (2007). HIV-1 Vpr function is mediated by interaction with the damage-specific DNA-binding protein DDB1. Proceedings of the National Academy of Sciences of the United States of America *104*, 4130-4135.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature *418*, 646-650.

Shibata, J., Yoshimura, K., Honda, A., Koito, A., Murakami, T., and Matsushita, S. (2007). Impact of V2 mutations for escape from a potent neutralizing anti-V3 monoclonal antibody during in vitro selection of a primary HIV-1 isolate. J Virol.

Simon, J.H., Fouchier, R.A., Southerling, T.E., Guerra, C.B., Grant, C.K., and Malim, M.H. (1997). The Vif and Gag proteins of human immunodeficiency virus type 1 colocalize in infected human T cells. J Virol *71*, 5259-5267.

Simon, J.H., Gaddis, N.C., Fouchier, R.A., and Malim, M.H. (1998). Evidence for a newly discovered cellular anti-HIV-1 phenotype. Nature medicine *4*, 1397-1400.

Simon, J.H., Southerling, T.E., Peterson, J.C., Meyer, B.E., and Malim, M.H. (1995). Complementation of vif-defective human immunodeficiency virus type 1 by primate, but not nonprimate, lentivirus vif genes. J Virol *69*, 4166-4172.

Sinclair, A.M., and Elliott, S. (2005). Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. J Pharm Sci *94*, 1626-1635.

Singh, I.R., Crowley, R.A., and Brown, P.O. (1997). High-resolution functional mapping of a cloned gene by genetic footprinting. Proceedings of the National Academy of Sciences of the United States of America *94*, 1304-1309.

Siwka, W., Schwinn, A., Baczko, K., Pardowitz, I., Mhalu, F., Shao, J., Rethwilm, A., and ter Meulen, V. (1994). vpu and env sequence variability of HIV-1 isolates from Tanzania. AIDS research and human retroviruses *10*, 1753-1754.

Slootstra, J.W., Kuperus, D., Pluckthun, A., and Meloen, R.H. (1997). Identification of new tag sequences with differential and selective recognition properties for the anti-FLAG monoclonal antibodies M1, M2 and M5. Mol Divers *2*, 156-164.

Smith, J.A., and Daniel, R. (2006). Following the path of the virus: the exploitation of host DNA repair mechanisms by retroviruses. ACS chemical biology *1*, 217-226.

Smith, W.A., Schurter, B.T., Wong-Staal, F., and David, M. (2004). Arginine methylation of RNA helicase a determines its subcellular localization. The Journal of biological chemistry *279*, 22795-22798.

Stellwagen, A.E., and Craig, N.L. (1997). Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7. Genetics *145*, 573-585.

Stelter, P., and Ulrich, H.D. (2003). Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. Nature *425*, 188-191.

Stone, J.C., Atkinson, T., Smith, M., and Pawson, T. (1984). Identification of functional regions in the transforming protein of Fujinami sarcoma virus by in-phase insertion mutagenesis. Cell *37*, 549-558.

Stremlau, M., Owens, C.M., Perron, M.J., Kiessling, M., Autissier, P., and Sodroski, J. (2004). The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature *427*, 848-853.

Strizki, J.M., Xu, S., Wagner, N.E., Wojcik, L., Liu, J., Hou, Y., Endres, M., Palani, A., Shapiro, S., Clader, J.W.*, et al.* (2001). SCH-C (SCH 351125), an orally bioavailable, small molecule antagonist of the chemokine receptor CCR5, is a potent inhibitor of HIV-1 infection in vitro and in vivo. Proceedings of the National Academy of Sciences of the United States of America *98*, 12718-12723.

Swingler, S., Gallay, P., Camaur, D., Song, J., Abo, A., and Trono, D. (1997). The Nef protein of human immunodeficiency virus type 1 enhances serine phosphorylation of the viral matrix. J Virol *71*, 4372-4377.

Tackett, A.J., DeGrasse, J.A., Sekedat, M.D., Oeffinger, M., Rout, M.P., and Chait, B.T. (2005). I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. Journal of proteome research *4*, 1752-1756.

Tang, H., Gaietta, G.M., Fischer, W.H., Ellisman, M.H., and Wong-Staal, F. (1997). A cellular cofactor for the constitutive transport element of type D retrovirus. Science *276*, 1412-1415.

Tang, H., McDonald, D., Middlesworth, T., Hope, T.J., and Wong-Staal, F. (1999). The carboxyl terminus of RNA helicase A contains a bidirectional nuclear transport domain. Molecular and cellular biology *19*, 3540-3550.

Tasaki, T., Mulder, L.C., Iwamatsu, A., Lee, M.J., Davydov, I.V., Varshavsky, A., Muesing, M., and Kwon, Y.T. (2005). A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. Molecular and cellular biology *25*, 7120-7136.

Tateishi, S., Sakuraba, Y., Masuyama, S., Inoue, H., and Yamaizumi, M. (2000). Dysfunction of human Rad18 results in defective postreplication repair and hypersensitivity to multiple mutagens. Proceedings of the National Academy of Sciences of the United States of America *97*, 7927-7932.

Tetsuka, T., Uranishi, H., Sanda, T., Asamitsu, K., Yang, J.P., Wong-Staal, F., and Okamoto, T. (2004). RNA helicase A interacts with nuclear factor kappaB p65 and functions as a transcriptional coactivator. European journal of biochemistry / FEBS *271*, 3741-3751.

Thomas, E.K., Connelly, R.J., Pennathur, S., Dubrovsky, L., Haffar, O.K., and Bukrinsky, M.I. (1996). Anti-idiotypic antibody to the V3 domain of gp120 binds to vimentin: a possible role of intermediate filaments in the early steps of HIV-1 infection cycle. Viral Immunol *9*, 73-87.

Topper, M., Luo, Y., Zhadina, M., Mohammed, K., Smith, L., and Muesing, M.A. (2006). Post-translational Acetylation of the HIV-1 Integrase Carboxyl-Terminal Domain is Dispensable for Viral Replication. J Virol.

Turelli, P., Doucas, V., Craig, E., Mangeat, B., Klages, N., Evans, R., Kalpana, G., and Trono, D. (2001). Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. Molecular cell *7*, 1245-1254.

Turlure, F., Devroe, E., Silver, P.A., and Engelman, A. (2004). Human cell proteins and human immunodeficiency virus DNA integration. Front Biosci *9*, 3187-3208.

Tyers, M., and Jorgensen, P. (2000). Proteolysis and the cell cycle: with this RING I do thee destroy. Current opinion in genetics & development *10*, 54-64.

Van Maele, B., Busschots, K., Vandekerckhove, L., Christ, F., and Debyser, Z. (2006). Cellular co-factors of HIV-1 integration. Trends in biochemical sciences *31*, 98-105.

Vanhecke, D., and Janitz, M. (2005). Functional genomics using high-throughput RNA interference. Drug discovery today *10*, 205-212.

Varshavsky, A. (1996). The N-end rule: functions, mysteries, uses. Proceedings of the National Academy of Sciences of the United States of America *93*, 12142-12149.

VerPlank, L., Bouamr, F., LaGrassa, T.J., Agresta, B., Kikonyogo, A., Leis, J., and Carter, C.A. (2001). Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). Proceedings of the National Academy of Sciences of the United States of America *98*, 7724-7729.

Vogt, V.M. (1996). Proteolytic processing and particle maturation. Current topics in microbiology and immunology *214*, 95-131.

Wain-Hobson, S., Vartanian, J.P., Henry, M., Chenciner, N., Cheynier, R., Delassus, S., Martins, L.P., Sala, M., Nugeyre, M.T., Guetard, D.*, et al.* (1991). LAV revisited: origins of the early HIV-1 isolates from Institut Pasteur. Science *252*, 961-965.

Wang, R., Cherukuri, P., and Luo, J. (2005). Activation of Stat3 sequence-specific DNA binding and transcription by p300/CREB-binding protein-mediated acetylation. The Journal of biological chemistry *280*, 11528-11534.

Weber, J., Rangel, H.R., Chakraborty, B., Tadele, M., Martinez, M.A., Martinez-Picado, J., Marotta, M.L., Mirza, M., Ruiz, L., Clotet, B.*, et al.* (2003). A novel TaqMan real-time PCR assay to estimate ex vivo human immunodeficiency virus type 1 fitness in the era of multi-target (pol and env) antiretroviral therapy. The Journal of general virology *84*, 2217-2228.

Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., Salazar-Gonzalez, J.F., Salazar, M.G., Kilby, J.M., Saag, M.S.*, et al.* (2003). Antibody neutralization and escape by HIV-1. Nature *422*, 307-312.

Wilchek, M., and Bayer, E.A. (1990). Methods in Enzymology (London, Academic Press).

Wiskerchen, M., and Muesing, M.A. (1995a). 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 *69*, 376-386.

Wiskerchen, M., and Muesing, M.A. (1995b). Identification and characterization of a temperature-sensitive mutant of human immunodeficiency virus type 1 by alanine scanning mutagenesis of the integrase gene. J Virol *69*, 597-601.

Wu, X., Liu, H., Xiao, H., Conway, J.A., Hehl, E., Kalpana, G.V., Prasad, V., and Kappes, J.C. (1999). Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions with the nucleoprotein reverse transcription complex. J Virol *73*, 2126-2135.

Wyatt, R., Kwong, P.D., Desjardins, E., Sweet, R.W., Robinson, J., Hendrickson, W.A., and Sodroski, J.G. (1998). The antigenic structure of the HIV gp120 envelope glycoprotein. Nature *393*, 705-711.

Xiao, Z., Ehrlich, E., Yu, Y., Luo, K., Wang, T., Tian, C., and Yu, X.F. (2006). Assembly of HIV-1 Vif-Cul5 E3 ubiquitin ligase through a novel zinc-binding domain-stabilized hydrophobic interface in Vif. Virology *349*, 290-299.

Yamashita, M., and Emerman, M. (2005). The cell cycle independence of HIV infections is not determined by known karyophilic viral elements. PLoS pathogens *1*, e18.

Yi, J., Arthur, J.W., Dunbrack, R.L., Jr., and Skalka, A.M. (2000). An inhibitory monoclonal antibody binds at the turn of the helix-turn-helix motif in the N-terminal domain of HIV-1 integrase. The Journal of biological chemistry *275*, 38739-38748.

Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P., and Yu, X.F. (2003). Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science *302*, 1056-1060.

Yu, Y., Xiao, Z., Ehrlich, E.S., Yu, X., and Yu, X.F. (2004). Selective assembly of HIV-1 Vif-Cul5-ElonginB-ElonginC E3 ubiquitin ligase complex through a novel SOCS box and upstream cysteines. Genes & development *18*, 2867-2872.

Yung, E., Sorin, M., Pal, A., Craig, E., Morozov, A., Delattre, O., Kappes, J., Ott, D., and Kalpana, G.V. (2001). Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1. Nature medicine *7*, 920-926.

Yung, E., Sorin, M., Wang, E.J., Perumal, S., Ott, D., and Kalpana, G.V. (2004). Specificity of interaction of INI1/hSNF5 with retroviral integrases and its functional significance. J Virol *78*, 2222-2231.

Zhang, H., Yang, B., Pomerantz, R.J., Zhang, C., Arunachalam, S.C., and Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature *424*, 94-98.

Zhang, S., and Grosse, F. (1997). Domain structure of human nuclear DNA helicase II (RNA helicase A). The Journal of biological chemistry *272*, 11487-11494.

Zheng, X.H., and Hughes, S.H. (1999). An avian sarcoma/leukosis virus-based gene trap vector for mammalian cells. J Virol *73*, 6946-6952.

Zhou, K., Choe, K.T., Zaidi, Z., Wang, Q., Mathews, M.B., and Lee, C.G. (2003). RNA helicase A interacts with dsDNA and topoisomerase IIalpha. Nucleic acids research *31*, 2253-2260.

Zhou, T., Xu, L., Dey, B., Hessell, A.J., Van Ryk, D., Xiang, S.H., Yang, X., Zhang, M.Y., Zwick, M.B., Arthos, J.*, et al.* (2007). Structural definition of a conserved neutralization epitope on HIV-1 gp120. Nature *445*, 732-737.