Functional Analysis On the Interactions of the Human Immunodeficiency Virus Type 1 Integrase With its Cofactors that Regulate Viral Replication

by

Yingfeng Zheng

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Medical Microbiology

University of Manitoba

Winnipeg

Abstract

Like all viruses, the replication of HIV-1 relies heavily on host proteins due to its limited genome products. HIV-1 integrase (IN) catalyzes the integration of viral DNA into host genome and also impacts other steps of viral replication cycle, all of which are assisted by various cellular proteins. Among them, LEDGF/p75 acts as the IN-to-chromatin tethering factor. However, whether other cellular cofactors also participate in this process still remains elusive. To gain insight into the mechanism of action of HIV-1 IN during viral integration, we used a previously described IN/yeast lethality system and our results revealed that the HIV-1 IN-induced yeast lethality absolutely required its chromatin binding ability. Since there is no yeast homolog of LEDGF/p75, it raises the possibility that IN may recruit other cellular cofactors for its chromatin targeting. Consistently, further analysis in mammalian cells indicated that HIV-1 IN was able to mediate chromatin binding independent of IN-LEDGF/p75 interaction and that HIV-1 fitness relied more on chromatin binding than LEDGF/p75 binding of IN. These data greatly enrich our current knowledge on the dynamic interplay within the ternary complex IN/LEDGF/chromatin.

HIV-1 exploits multiple cellular cofactors not only to facilitate viral replication, but also to evade the host defense system in favor of the virus. IN is known to be an unstable protein, degraded by the host ubiquitin-proteasome pathway. To investigate how IN avoids the host degradation machinery in the context of viral infection, we showed that IN interacted with host protein Ku70 and protected itself from the Lys48-linked polyubiquitination proteasomal pathway. More importantly, Ku70 was shown to be incorporated into the progeny virus in an IN-dependent manner, and both cell- and virus-

associated Ku70 were essential for HIV-1 replication. Finally, the data demonstrated that the interactions between HIV-1 IN and host cofactors can be regulated through its SUMO-interacting motifs (SIMs). Three putative SIMs (72VILV75; 200IVDI203 and 257IKII260) in IN were examined and shown to be essential for IN-LEDGF/p75 but not IN-Ku70 interaction.

In summary, this study advances our knowledge of the interaction network between IN and its cofactors, which would have important implications for the design of anti-HIV drugs.

Acknowledgements

To begin with, I would like to express many thanks to my supervisor, Dr. Yao, for his enthusiasm for research and teaching, his patience at explaining all the questions I have. I very much appreciate everything he has done in the course of the supervision of this thesis. I thank him for his kindness and for the incredible encouragement to pursue this research.

I present my warmest gratefulness to my graduate studies advisory committee members Dr. Xi Yang, Dr. Jude Uzonna and Dr. Tooru Mizuno, for their moral support, mentoring and very useful comments in the past a few years. Their encouragement and advice have greatly helped to bolster my confidence. It is dedicated mentors like them who make a difference in my life. I sincerely appreciate my external examiner Dr. Heinrich G. Göttlinger for his time, effort, and feedback on my thesis.

I must thank all current and former members of Dr Yao's laboratory. I would like to give special thanks to Dr. Zhujun Ao for her support and guidance throughout this entire process. Much appreciation is also due to Zaikun Xu, Kallesh Danappa Jayappa, Xiaoxia Wang, Kamal Parvez, Zhe Yu, Dr. Yuanyuan Gu, Jinyu Peng, Ranzhun Zhao, Dr. Rong Zhu and Dr. Liyu Chen for their support, assistance and important discussions.

I would also like to take this opportunity to thank everybody in the Department of Medical Microbiology for their support, collaboration and alliance during my study. Especially, I am indebted to Angela Nelson, Sharon Tardi, John Rutherford, Sue Ramdahin, Cheryl Reimer for their guidance and technical support. I would like to thank

all mentors, staffs and trainees in CIHR International Infectious Disease and Global Health training program (IID & GHTP) for their exceptional coaching and precious discussions. Special thanks go to Dr. Keith Fowke, Jude Zeiske and Eva Lindsey.

My thanks are also directed at many people who have contributed reagents and materials for this thesis. I thank in particular Martin Clement and Dr. Pierre Belhumeur from Université de Montréal for their technical support on yeast chromatin binding assay; Dr. Dhirendra P. Singh from University of Nebraska Medical Center for providing EGFP-LEDGF/p75 plasmid; Dr. Heinrich G. Göttlinger (UMass Medical School) for HA-Ub plamids, Dr. Shigemi Matsuyama from Case Western Reserve University for Ku70 knockout mouse embryonic fibroblasts and Dr. Ronald T Hay for V5-Ubc9 and HA-SUMO1 plasmids (University of Dundee, UK).

I truly appreciate the assistance and encouragement which I received from my family near and far. I wish to take this chance to express my gratitude to my parents for believing in me all the way long. In addition, I do not have the words to articulate my love and thankfulness to my husband, Sherwin, for his patience, endurance and for proof reading this thesis. Most notably, I do want to thank my beloved little daughter Jamalia who has made my life richer and more meaningful.

Finally, I am grateful to MHRC/MICH and IID & GHTP for the generous support which made this work possible.

Dedication

This dissertation is dedicated to my family for their love and support.

Lists of tables

Lists of Figures

Abbreviations

2-LTR	two-long terminal repeat
3-D	three-dimensional
Aa	amino acid
AIDS	Acquired immunodeficiency syndrome
APOBEC3G	apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G
ARV	antiretroviral
ASLV	avian sarcoma-leukosis virus
Atr	Atm and Rad related
BAF	barrier to auto-integration factor
CA	Capsid
CBDs	chromatin binding domains
CCD	catalytic core domain
Co-IP	co-immunoprecipitation
CRs	charged regions
CTD	C-terminal domain
DAPI	4'-6-Diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagles Medium
DNA-PKcs	catalytic subunit of DNA-dependent protein kinase
DUB	Deubiquitination
ECL	Enhanced chemiluminescence
EED	*Polycomb* group embryonic ectoderm development
EIAV	equine infectious anemia virus
ER	endoplasmic reticulum
ERK2	Extracellular signal-regulated kinase 2
EVG	Elvitegravir

FCS	fetal calf serum
FDA	Food and Drug Administration
FITC	Fluorescein isothiocyanate
FIV	feline immunodeficiency virus
GFP	Green Fluorescent Protein
HAART	Highly Active Antiretroviral Therapy
HATs	histone acetyltransferases
HDAC	histone deacetylase
HDGF	hepatoma-derived growth factor
HIV	Human Immunodeficiency Virus
HMGA1	high-mobility group protein 1
hRad18	Human Rad18
HRP	horseradish peroxidase
HRPs	HDGF-related proteins
Hsp	Heat shock protein
HTH	helix-turn-helix
HTLV	human T-cell leukaemia virus
IBD	IN binding domain
Impα3	Importin α3
Imp7	Importin 7
Impβ	Importin β
IN	integrase
Ini1	Integrase interactor 1
IP	Immunoprecipitation
JNK	c-Jun N-terminal kinase
kD	Kilodalton
KD	Knockdown
LEDGF	Lens epithelium-derived growth factor

LTR	long terminal repeat
MA	Matrix
MHCI	major histocompatibility class I
M-MLV	moloney murine leukemia virus
MOI	multiplicity of infection
NC	Nucleocapsid
NC	negative control
NDR	Nuclear DBF-2-related
Nef	negative factor
NHEJ	non-homologous end joining
NLS	nuclear localization signal
NMR	nuclear magnetic resonance
NNRTIs	non-nucleoside reverse-transcriptase inhibitors
NPCs	nuclear pore complexes
NRTIs	reverse transcriptase inhibitors
NTD	N-terminal zinc-binding domain
Nups	Nuclearporins
PBS	Phosphate buffered saline
PCNA	proliferating cell nuclear antigen
PIC	Preintegration Complex
Pin1	peptidyl prolyl-isomerase enzyme
PKA	Protein kinase A
PPIs	protein – protein interactions
PR	Protease
PTM	post-translational modification
PWWP	Pro-Try-Try-Pro
qPCR	quantitative PCR
RAL	Raltegravir

RANBP 2	RAN binding protein 2
RLU	relative light units
RRE	Rev response element
RT	reverse transcriptase
RTC	reverse transcription complex
S. cerevisiae	Saccharomyces cerevisiae
SAP	SAF-A/B, Acinus and PIAS
SD	standard deviation
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SENP	sentrin-specific proteases
shRNAs	short hairpin RNAs
SIM	SUMO-interacting motif
siRNAs	small interfering RNAs
SIV	Simian immunodeficiency virus
STI	strand transfer inhibitor
SUMO	small ubiquitin-like modifier
SUMO-CS	SUMOylation consensus sites
TAR	Trans-activation Response
Tat	Trans-Activator of Transcription
TNPO3	Tranportin 3
Ub	Ubiquitin
Ubls	ubiquitin-like proteins
UBR	ubiquitin ligase N-recognin
UPS	ubiquitin proteasome system
VBP1	von Hippel–Lindau binding protein 1
VHL	von Hippel-Lindau protein
Vif	Virus Infectivity Factor

Vpr	Viral Protein R
VSV-G	vesicular stomatitis virus G glycoprotein
vWA	von Willibrand A domain

TABLE OF CONTENT

Chapter 1 Introduction

In this chapter, the epidemiology of Human Immunodeficiency Virus (HIV), its genome organization, viral structure, HIV-1 replication life cycle will be discussed. Then, I will review the structure and functions of HIV-1 integrase (IN) and its cellular cofactors, post-translational modifications (PTMs) of IN in the context of retroviral pathogenesis and discuss current efforts on IN as the new target for antiretroviral (ARV) therapy.

1.1 HIV-1 epidemiology

HIV is the pathogen which is responsible for Acquired immunodeficiency syndrome/AIDS. HIV belongs to the family Retroviridae genus lentiviruses, or 'slow' viruses. HIV, the virus itself does not cause death in human beings; instead it suppresses normal immunity and causes opportunistic infections and cancers. Since the discovery of HIV/AIDS in 1983 [1, 2], its pandemic has posed a great risk to the well-being of humans and the whole society. According to 2011 UNAIDS (the Joint United Nations Programme on HIV/AIDS) World AIDS Day report, there are still 34 million people worldwide living with HIV. Amongst this population, there are around 2.7 million new HIV infections and 1.8 million people died of AIDS-related illnesses in 2010 [3]. Currently, there are two different circulating virus strains of HIV including HIV-1 and HIV-2, with HIV-1 more virulent and predominant. While HIV-1 is worldwide, HIV-2 cases are primarily found in West Africa with relatively low prevalence [4]. HIV-1 strains are categorized into three clades: M (Major/Main), N (Non-M, Non-O/New) and O (Outlier). And each group is further divided into various subtypes [5]. HIV-1 M subtype is the most

prevailing virus strain responsible for the worldwide infection. In this thesis, unless specified otherwise, HIV refers to HIV-1.

1.2 HIV-1 genome, structure and viral replication cycle

1.2.1 Organization of HIV-1 virion

HIV is an enveloped, nearly rounded virus with a diameter of 120nm (Figure 1A). The virus particles are surrounded by a lipid envelope derived from the host cell plasma membrane. The viral exterior envelope glycoproteins gp120 and gp41 are organized into trimetric complexes which are embedded within the lipid bilayer [6]. The viral core (or Capsid) is usually cone-shaped and made from the viral capsule protein p24. Inside the core, there are two single strands of viral RNA, together with three enzymes required for HIV replication called reverse transcriptase (RT), integrase (IN) and protease (PR). HIV's core also includes nucleocapsid (NC) protein or p7 and three accessory proteins Nef, Vif and Vpr [7]. Lying between the viral core and the envelope are HIV matrix (MA) proteins. HIV-1 also incorporates a wide variety of cellular proteins including cellular chaperones Cyclophilin A, Heat shock proteins (Hsps) including Hsp40, Hsp60, Hsp90 and Hsp70; cellular trafficking systems including actin, tubulin; nuclear proteins histone deacetylase 1 (HDAC 1) and the chromatin remodeling protein Integrase interactor 1 (Ini1)/HSNF5, cytidine deaminase APOBEC3G (apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G) protein and kinases Extracellular signal-regulated kinase 2 (ERK2), Protein kinase A (PKA) and Nuclear DBF-2-related (NDR) kinases 1/2 (see a review [8]).

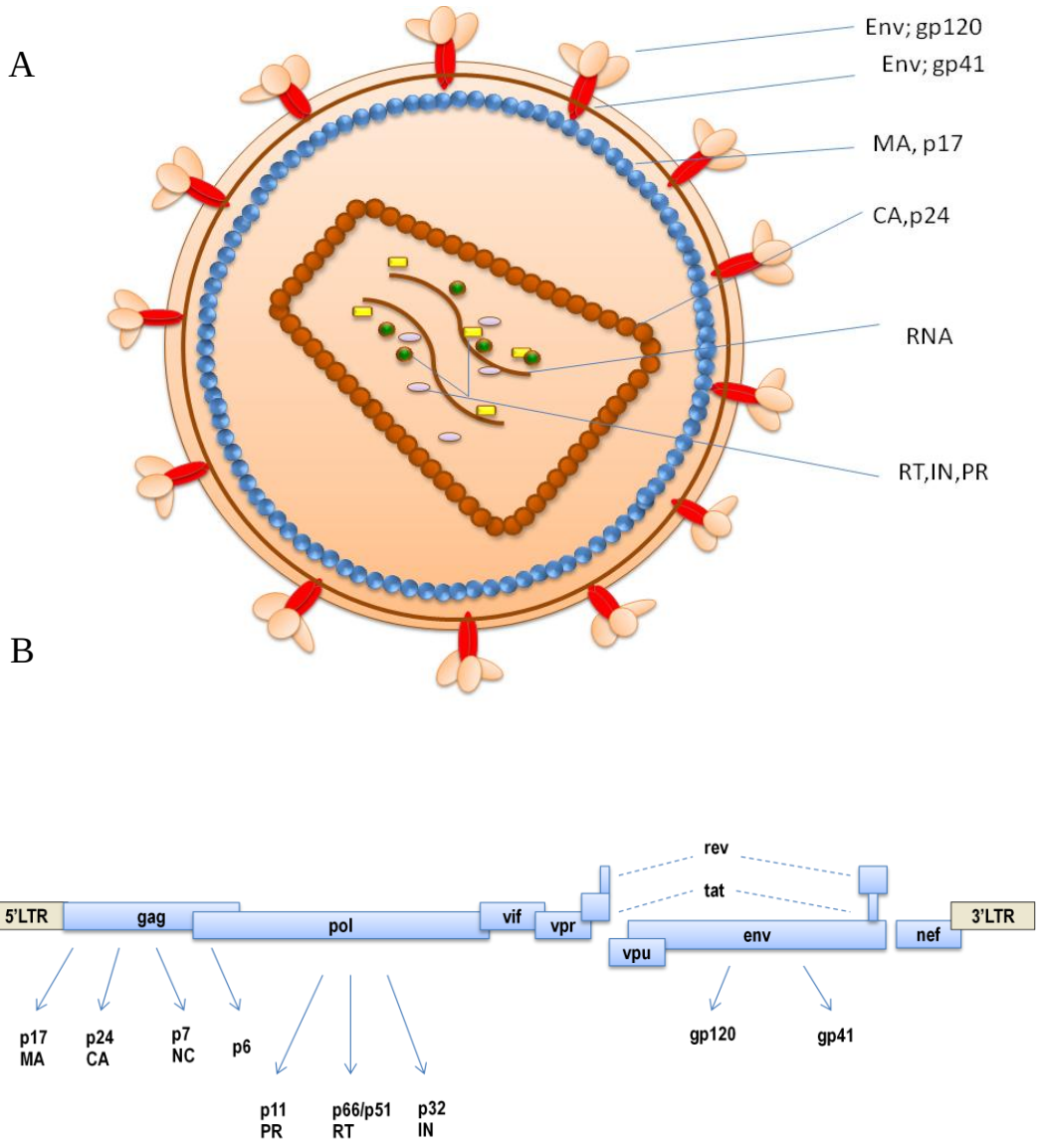

Figure 1 Schematic diagram of HIV virion and genome. A) The outer envelop proteins gp120 and gp41 are embedded within the lipid bilayer. Viral protein p17 forms matrix while an inner layer of viral protein called p24 forms the core. Inside the cone-shaped viral core, there are two copies of the single stranded RNA and three viral enzymes RT, PR and IN. B) A schematic representation of the HIV-1 gene products encoded by the HIV-1 ~10kb genomic sequence. The HIV-1 genome contains nine open

reading frames which produce 15 proteins and has a LTR at each end of its genome. RT: reverse transcriptase; PR: protease; IN: integrase. LTR: long terminal repeat.

1.2.2 Structure of the viral genome

HIV-1 is an enveloped RNA virus that replicate through a DNA intermediate. Each HIV-1 virion contains two copies of single-stranded RNA genome, which is around ~10kb long [9, 10]. The genome not only carries the genetic information of the virus, but also harbors internal structures fundamental to viral replication, including Trans-activation Response (TAR) hairpin, the packaging signal, the dimerization site, the Gag-Pol frameshifting region, the Rev Response Element (RRE), etc. These elements have been shown to play various roles in the viral replication cycle such as transcription activation, reverse transcription initiation, genomic dimerization, frameshifting reading frames, RNA nuclear export, signal polyadenylation, and interaction with viral and host proteins [11-13]. The HIV-1 genome contains nine open reading frames which produce 15 proteins (Figure 1B) [11]. gag and env genes encode structural proteins to form new viral particles. gag gene encodes Gag precursor protein which is synthesized and proteolytically cleaved into MA, CA (capsid), nucleocapsid (NC), and p6. pol gene encodes three viral enzymes: PR, RT and IN. Other six genes tat, rev, nef, vif, vpr and vpu code two regulatory Proteins Tat and Rev, and four accessory proteins Nef, Vpr, Vpu and Vif. Among these, Vif, Vpr and Nef are incorporated into new viral particles. In HIV-2, vpu gene is replaced by vpx. These proteins termed "accessory" are due to the initial finding that they were dispensable for viral replication in *ex vivo* cellular systems. However, recent studies suggest that HIV-1 develops those accessory proteins such as Vif, Vpr and Vpu to counteract the cell-mediated antiviral defenses [14-16]. A brief description on the functions of each HIV-1 encoded protein is discussed below.

Gag gag gene encodes 55-kilodalton (kD) Gag precursor protein or p55 from the unspliced viral mRNA. Gag precursor protein plays key roles in the viral assembly and budding steps. After Gag is synthesized on cytosolic ribosomes, the N terminus of Gag is first myristoylated and then targeted to the cytoplasmic side of cell membranes [17], where Gag-Gag multimers assemble into higher order structure. Meanwhile, Gag binds two copies of the viral RNA and other viral proteins such as Gag-Pol precursor protein and cellular proteins, and coordinates the incorporation of viral and host components into the assembling particle. After budding, the virus is still immature or non-infectious till processing of precursor proteins takes place. During maturation, Gag precursor protein is cleaved by viral PR into four smaller proteins designated MA (p17), CA (p24), NC (p9), p6 and two spacer peptides SP1 and SP2 [18].

The MA protein is derived from the N-terminal, myristoylated end of p55. Most of MA proteins lie beneath the inner surface of the virion lipid bilayer to ensure structural support. A small amounts of MA localize inside the core and are found in tight association with the HIV-1 core and the Preintegration Complex (PIC) [19]. MA plays various roles in both early and late stages of viral replication cycle. The most important role that MA functions is to mediate the plasma membrane targeting of Gag polyproteins. Meanwhile, as a component of PIC, MA facilitates the nuclear transport of the viral genome, which allows HIV infect nondividing cells [20, 21]. Recent studies have also highlighted that MA has the immunoregulatory effects during HIV-1 infection by binding to cellular receptor p17R. Therefore, MA promotes proliferation of T cells and release of proinflammatory cytokines, migration of immune cells, etc. [22-25].

CA forms a closed conical-shaped capsid shell that encapsulates the viral RNA, NC proteins, and three viral enzymes in the mature, infectious virus [7]. As a major structural protein, CA also carries non-structural functions, affecting both the early and late steps of viral replication. Improper self-assembly of CA into cones during viral assembly or abnormal disassembly to release genetic contents into cytoplasm have been shown to impair viral infectivity [26-28].

The NC domain of Gag is involved in the assembly step such as Gag-Gag interaction and Gag binding with cellular components in close proximity to the cellular membrane, and packaging of the RNA genome into the new viral particle (see review [29]). NC protein might not be strictly required for reverse transcription and integration during the early stage of viral replication. But in the mature virion, the presence of NC might serve as a nucleic acid chaperone which enhances reverse transcription and integration ([30-32] or see review [29]).

The 52-amino acid (52-aa) p6 protein is the smallest viral protein among all the lentiviruses [33]. Although small in size, p6 plays pivotal roles in the HIV-1 replication cycle. p6 domain of Gag is important for binding and incorporation of viral proteins such as Vpr, Pol and Env proteins into assembly virions, as well as recruiting cellular components from the human multivesicular bodies pathway to facilitate budding of virions [33].

gag-pol Ribosomal frameshifting at the gag-pol junction (with a frequency of 5 to 10%) results in Gag-Pol polyprotein precursor. pol encodes three viral enzymes PR, RT and IN. PR cleaves Gag and Gag-Pol polyproteins to produce the final MA, CA, NC, and p6

proteins from Gag and PR, RT, and IN proteins from Pol. Before the viral genome can be integrated into the host chromosome, the single-stranded viral RNA has to be converted into double-stranded DNA by RT. RT is a heterodimer comprised of a p66 and a p51 subunit with p66 subunit containing both the polymerase and RNase H activities, while p51 provides structural support [34]. HIV PR belongs to the family of aspartic proteases, and it recognizes the different substrate cleavage site sequences and cleaves the Gag and Gag-Pol polyproteins which are accumulated at the plasma membrane during or shortly after the release of assembled virus particles [35, 36]. The detailed description of HIV-1 IN will be discussed below.

Env Env precursor protein gp160, encoded by env, is cleaved post-translationally into the extracellular subunit gp120 and the transmembrane gp40, performed by furin-like cellular proteases [37]. Env (gp160) is synthersized in the endoplasmic reticulum (ER) and glycosylated in the ER and Golgi apparatus. The precursor is then cleaved to produce the non-covalently associated gp40/gp120 trimeric glycoprotein complex, which is then transported to the cell membrane for virus assembly [7]. The viral entry is accomplished through the interaction between Env proteins gp120/gp40 and host cell surface receptors and chemokine coreceptors (CD4 and CCR5/CXCR4) in the immune cells.

Two regulatory proteins

Tat Tat or (Trans-Activator of Transcription) is a relatively small viral protein, containing 86–101 residues depending on different viral strains. The most profound function of the Tat protein is enhancing viral transcription through three distinct mechanisms: binding the specific sequences of TAR located in the 5' end of LTR,

modulating activity of polymerase II and promoting NF-κB activation [38]. In addition, Tat is proposed to act as a viral toxin which can be secreted from infected cells and endocytosed into various cells types including endothelial cells, lymphocytes, monocytes, macrophages, neurons, thereby affecting gene expressions, cytokine productions and its related biological activities [38, 39].

Rev Rev refers to regulator of expression of virion. Rev is known to affect the late stage of viral replication cycle by binding to the RRE. Rev enhances the expression of the viral structural proteins Gag, Pol and Env by targeting viral RNA to the cellular CRM-1-mediated nuclear-export pathway, leading to increased cytoplasmic viral RNAs and increased expression of the encoded proteins [38]. However, recent studies have highlighted the novel functions of Rev in the early stage of viral replication in which Rev inhibits or regulates integration events [40].

Four accessory proteins

Vif Virus Infectivity Factor (Vif) is a 192-aa cytoplasmic protein [41]. Vif is known to counteract host anti-viral factor APOBEC3G by binding and targeting APOBEC3G to host proteasomal degradation [42, 43]. APOBEC3G acts to deaminate cytidines to uridines in the single-stranded DNA, thereby generating stop codons or G-A mutation in the newly synthesized viral cDNA [44, 45]. Vif-deleted virions are defective in the reverse transcription step, and Vif has been shown to regulate PR activity and the proteolytic processing of the Gag precursor [46-48].

Vpr Viral Protein R (Vpr) is a 96-aa small basic protein. The Vpr protein is ubiquitous in HIV-1 infected patients, present in virions, cells, sera and cerebrospinal fluid, etc.

Though small, Vpr has been demonstrated to have multiple functions during viral replication, such as the nuclear import of the PIC, cell cycle G2/M progression, apoptosis, and transactivation of the HIV-1 LTR and host cell genes (see a review [38]).

Vpu Vpu is an 81-aa dimeric integral membrane protein [49]. The biological functions of Vpu include induction of CD4 degradation and enhancement of viral particle release, promotion of apoptosis through inhibition of NF-κB [49-52].

Nef HIV-1 negative factor (Nef) is a 206-aa, N-terminally myristoylated regulatory factor. Nef has long been recognized to be related with disease progression [53, 54]. Nef has a few distinct roles during viral pathogenesis as revealed by *in vitro* infection model: downregulating CD4 receptors and cell surface major histocompatibility class I (MHCI) molecules, and modulating cellular signaling and activation, enhancing viral infectivity via CD4-independent mechanisms [55].

1.2.3 An overview of HIV-1 replication cycle

The HIV replication cycle includes nine steps: 1) binding and entry; 2) uncoating; 3) reverse transcription; 4) nuclear import; 5) provirus integration; 6) transcription of viral RNA; 7) virus protein synthesis; 8) assembly budding and 9) maturation (Figure 2). The entire life cycle of HIV-1 is arbitrarily divided into the early and the late stage by integration. The early stage begins with binding till integration, whereas the late stage refers to the steps between transcription of viral genome to the maturation step. Like many other intracellular pathogens, HIV extensively exploits the host to aid in its replication or to bypass the immune responses and counteract host anti-viral restriction factors. Almost all the steps of HIV-1 replication involve the intimate interplay between

viral proteins with a wide range of cellular proteins and or cellular pathways. As such, HIV-1 replication is the result of co-evolution between the virus and the host. For more detailed HIV-1 replication implicated with multiple host factors, please refer to reviews [56-58]. Here, only a concise summary on HIV-1 life cycle is provided below.

The **entry** step can be summarized as the binding and fusion of the viral envelop with the plasma membrane of the host cell. Binding of gp120 with CD4 receptors on the surface of target cells as well as coreceptors such as CXCR4 or CCR5 leads to conformational changes of gp41 followed by the insertion of the gp41 fusion peptide into the target membrane, resulting in the fusion of viral and cellular membranes and the release of the viral core into the cytoplasm [59]. After binding and entry, uncoating occurs. **Uncoating** is the process that intact core consisted of CA proteins disassembles to release genomic RNA into cytoplasm. It has been proposed that phosphorylation of CA induces the dissociation of CA-CA protein interactions and promotes the disassembly of the HIV-1 core, the process which is assisted by cellular factors [60, 61]. Following uncoating, core components are then reconstructed into a complex named reverse transcription complex (RTC), within which the viral RNA is converted into double-stranded DNA catalyzed by viral RT named **reverse transcription**. This process is accomplished through three catalytic functions by viral enzyme RT: RNA-directed DNA synthesis, DNA-directed DNA synthesis and DNA-directed RNA hydrolysis [62]. Reverse transcription takes place mainly in the cytoplasm, while it also occurs as early as in the free virion before viral entry, with a small portion persisting even after nuclear import [63, 64]. After the completion of reverse transcription, RTC is now termed PIC, as it has been shown to be integration-competent and capable of integrating into DNA targets *in vitro* [65, 66]. PIC

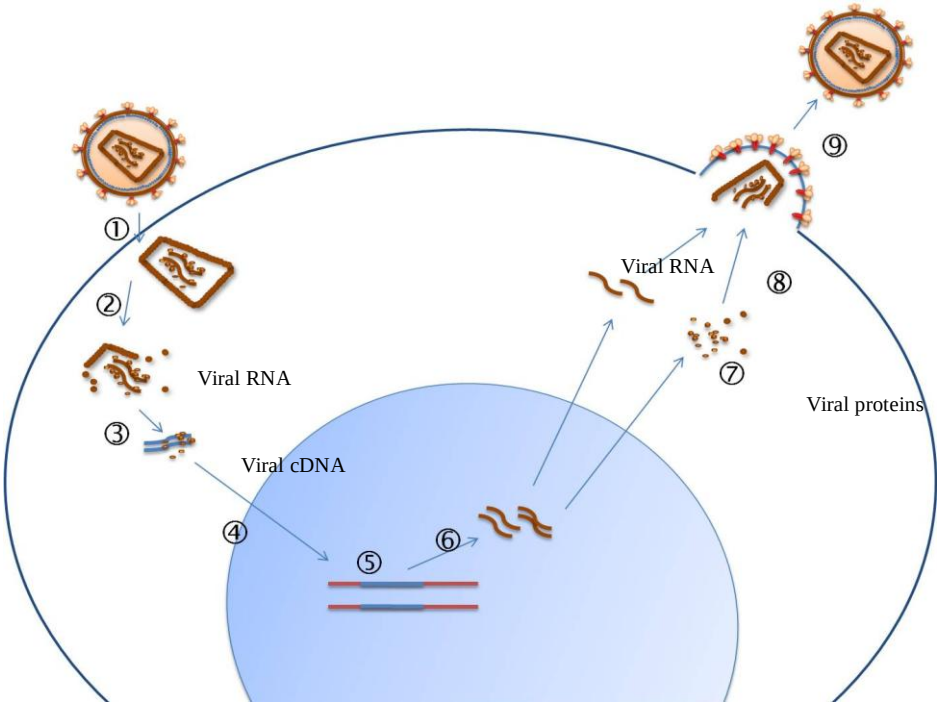

Figure 2 Schematic of the HIV-1 replication cycle. Nine important steps of the replication cycle are numbered: 1) binding and entry; 2) uncoating; 3) reverse transcription; 4) nuclear import; 5) provirus integration; 6) transcription of viral RNA; 7) virus protein synthesis; 8) assembly budding and 9) maturation.

formed in the host cellular cytoplasm is a functional nucleoprotein complex in which newly reverse transcribed viral DNAs form a complex with both viral proteins, including IN, MA, NC, RT, Vpr and various cellular proteins (reviewed by Al-Mawsawi LQ et al.) [67]. These cellular proteins include Lens epithelium-derived growth factor (LEDGF), Ini1, high-mobility group protein 1 (HMGA1), barrier to auto-integration factor (BAF), Hsp60, *Polycomb* group embryonic ectoderm development (EED) protein, etc. (for a review see [68]). In order for HIV-1 reverse transcribed viral DNA integration into host genome, PIC has to pass across nuclear membrane. **Nuclear import** of PIC is achieved through two different mechanisms: entry through breakdown of the nuclear membrane during mitosis in non-dividing cells and active nuclear translocation pathway in both dividing and non-dividing cells [21]. After PIC reaches the nucleus, HIV-1 IN catalyzes the **integration** of its viral DNA genome into the host cell chromosome. Thereafter, virus is able to utilize the cellular replication machinery for its genome transcription and subsequent translation. The integrated viral DNA is known as provirus. The detailed description of nuclear import and integration will be discussed in the next section.

The late stage of HIV-1 replication starts from **transcription,** the synthesis of the viral RNAs from integrated provirus. The HIV-1 transcription begins from U3 promoter within the LTR with Tat protein greatly increasing LTR-driven transcription. Transcription results in three different viral RNAs: non-spliced transcript which encodes HIV structural proteins (Gag, Gag–Pol precursors) and serves as genomic RNA packaged into the new virion; partially spliced and multiply spliced mRNA [69]. In the early phase, multiply spliced mRNA is easily exported to the cytoplasm and translated into Rev, Tat and Nef, while in the late phase unspliced and partialy spliced RNAs which produce Gag, Gag-Pol,

Env, Vif, Vpr and Vpu are exported to the cytoplasm through Rev [69]. HIV-1 Rev binds to the RRE, an cis-acting RNA element present in all unspliced and partially spliced HIV-1 RNAs, and Rev/RRE complex are transported into the cytoplasm via interaction with the cellular nuclear export machinery [69]. In addition to transcription from integrated DNA, non-integrated DNA can also serve as a template to generate Tat and Nef (and potentially Rev) transcripts [70]. Once in the cytoplasm, viral mRNAs are **translated** by utilizing cellular translational machinery [71]. Studies have shown that HIV-1 Gag is sufficient to assemble virus-like particles at the plasma membrane and to bud off the cells [72]. Due to its N‑terminal myristic acid and conserved basic amino acid residues [17, 73, 74], the MA domain of Gag initiates **viral assembly** by targeting and binding the plasma membrane so that the Gag and Gag-Pol polyproteins are localized to the cell membrane. Meanwhile, Env is transported to the cell surface, coating the new particle with gp120 and gp41, for which HIV-1 utilizes Vpu and Nef to counteract the presence of CD4: Vpu promotes CD4 degradation in the ER where both Env and CD4 are synthesized, and Nef promotes endocytosis and degradation of cell surface CD4 [11]. The viral core is assembled from the Gag and Gag-Pol polyproteins, Vif, Vpr, Nef, and the genomic RNA, and the immature virion begins to **bud** from the cell surface. Concomitant with or soon after budding, **maturation** begins, in which Gag and Gag-Pol polyproteins are proteolytically processed by PR to yield MA, CA and NC to construct mature viral cores and viral enzymes RT, PR and IN [75]. The mature and infectious virion is then ready to infect new cells.

1.3 HIV-1 IN

This section, while reiterating certain points discussed above, will mainly focus on structure and functions of HIV-1 IN, its cellular cofactors, PTMs of IN and current anti-IN strategies in the drug development.

1.3.1 Domain structure of IN

As discussed above, HIV-1 IN is encoded by the pol gene. IN is first synthesized as part of Gag-Pol polyprotein, in which Pol is cleaved into three viral enzymes RT, PR and IN during maturation. IN is a 288-aa, 32kDa viral protein, contains three distinct structural domains, the N-terminal zinc-binding domain (NTD, residues 1 – 49), the central catalytic core domain (CCD, residues 50 – 212) and the C-terminal domain (CTD, residues 213 – 288) (Figure 3A). The NTD contains a conserved His-His-Cys-Cys or HHCC motif which is responsible for binding Zn2+ atom and promoting oligomerization [76, 77]. Subsequent studies suggested that both of CCD and CTD domains of IN are required for its multimerization while NTD is not important [76, 78, 79]. Retroviral INs belong to DDE(D) nucleotidyltransferases which also include RNaseH as well as MuA and Tn5 transposases. The IN CCD contains three highly conserved residues Asp 64, Asp 116 and Glu 152 (the DDE motif) which coordinate a divalent cation (Mg2+ or Mn2+), essential for the enzyme's catalytic activity [80]. Mutations introduced to any of these residues abolish all catalytic activities of IN including 3’ processing, 5’ strand transfer and disintegration [81, 82]. The CTD contains an SH3-like fold and its main function is to bind DNA specifically or non-specifically [83-85]. CTD is also of importance in the binding of a variety of viral and host cellular proteins, such as RT, Importin α3 (Impα3), p300, Ku70 [86-88]. IN functions as a dimer or even higher order

oligomers. The dynamic subunit-subunit interaction of IN is essential for the biological functions of IN [89]. The HIV-1 IN dimer crystal structure is shown in Figure 3B.

1.3.2 Functions of IN

Functions of IN may be divided into two major categories: integration and non-catalytic activities of IN, the latter include any other effects of IN in the viral replication including reverse transcription, nuclear import, transcription, gene expression, and assembly. Here, all these functions will be discussed in the order of occurrence during the viral life cycle.

1.3.2.1 Reverse transcription

IN is a pleiotropic protein that affects different steps throughout the viral life cycle including reverse transcription, PIC nuclear import, integration and post-integration steps, such as viral protein expression, transcription, packaging and processing [16, 90-92]. The effect of IN on reverse transcription has been extensively observed in the past [90, 91, 93, 94], which is attributed to the direct interaction between RT and IN [88, 92]. Further studies revealed that both subunits of RT, p51 and p66, are able to bind to HIV-1 IN, and the binding domain was mapped to the CTD of IN [88, 95]. Two possibilities might account for the effect of their interaction on reverse transcription. The interaction might change the architecture of the RTC where reverse transcription takes place; it is also proposed that IN binding with cellular protein SIP1/Gemin2 enhances reverse transcription through facilitating the assembly of RT on viral RNA or increased binding affinity of RT with RNA strands [92, 96, 97].

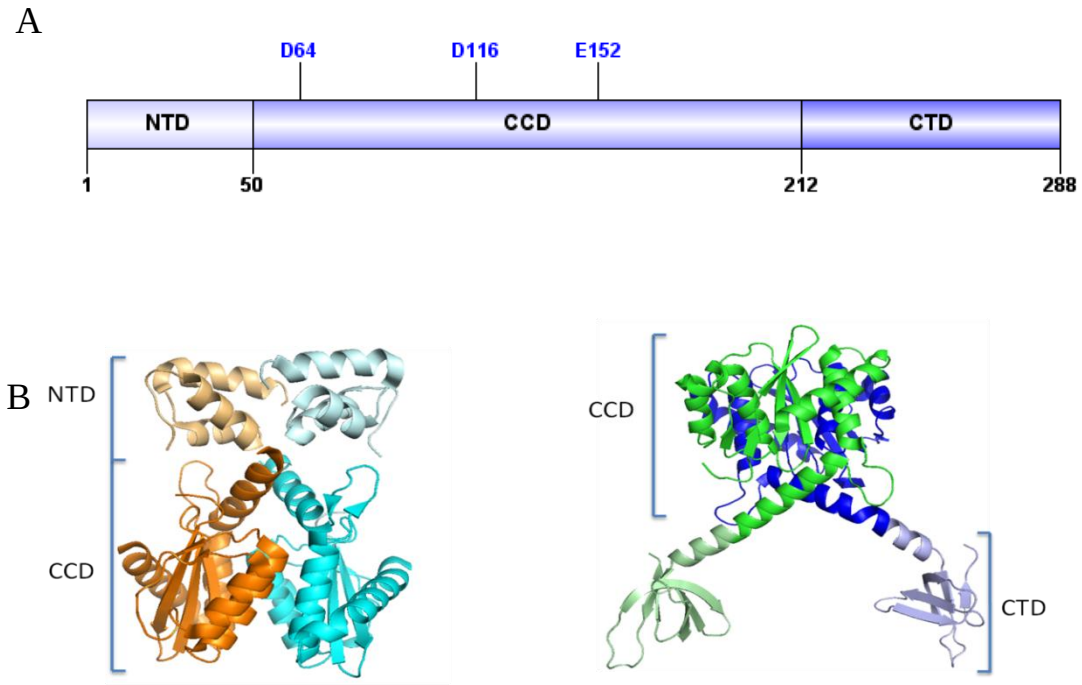

Figure 3 Schematic diagram of HIV-1 IN and its dimer crystal structure. A) HIV-1 IN is a 288-aa viral protein with three structurally distinct domains: NTD, CCD and CTD. The CCD contains the highly conserved DDE (D64/D116/E152) motif. B) Ribbon representation of dimeric IN fragments containing NTD plus CCD (left) and CCD plus CTD (right). Left: the colors used for each subunit are: palecyan for NTD and cyan for CCD in one subunit, while the other subunit NTD is shown as light orange and CCD is shown as orange color. Right: CCD and CTD from one subunit are shown as blue and light blue respectively while the other are shown as green and light green. The structures were derived from PDB files 1K6Y [98] and 1EX4 [99] respectively. (All these structures were drawn in PyMol from http://www.pymol.org.) NTD: N-terminal domain; CCD: catalytic core domain; CTD: C-terminal domain.

1.3.2.2 Nuclear import

The major hallmark between lentiviruses such as HIV-1 and gammaretroviruses such as the moloney murine leukemia virus (M-MLV) lies in the fact that the former can infect non-dividing cells as well as dividing cells at interphase while M-MLV has to wait for the nuclear envelope to break down during mitosis to enter the nucleus of their target cells for successful replication (see reviews [100, 101]). HIV-1 shows no dependence on the cell cycle progression [102] due to the capability of HIV-1 PIC, which actively transports into the nucleus through the nuclear pore complexes (NPCs) with the help of host nuclear import machinery. PIC, a large multiprotein complex consisted of both viral and host components, is estimated to be more than 50 nm in diameter, while NPC only allows molecules with a diameter up to 9 nm or smaller than 40–45 kDa in molecular weight to passively transport between the cytoplasm and the nucleoplasm (see a review [101]). Thus, passive diffusion across the nuclear pore is impossible for HIV-1 PIC. Indeed, active transportation of viral PIC might be beneficial for infection and might contribute to a rapid replication rate in the infected patients [103]. The active nuclear import step of HIV-1 PIC has been attributed to the karyophilic viral proteins MA, Vpr, IN within PIC and cis-acting central DNA flap, but the exact mechanisms of nuclear entry still remain unclear [21]. The karyophilic viral factors MA, Vpr and IN contain classical or atypical nuclear localization signal (NLS) that facilitates the nuclear import of HIV-1 PICs [21, 80, 104, 105]. However, their contributions to the nuclear import still remain unclear and questionable since mutations within all these identified NLSs were still replicable in non-dividing cells and additional NLS to gammaretroviruses did not confer the ability to infect non-dividing cells [106-108]. Another HIV-1 viral protein CA

has also been implicated in the nuclear import of viral DNA, although CA is not nucleophilic [100]. However, following studies suggest that uncoating of CA from the nucleoprotein core is a prerequisite to expose the karyophilic elements within the PICs, which in turn impacts on nuclear import of viral PIC [100, 101]. Nonetheless, it seems reasonable to consider that all these viral proteins are important but not the sole determinants responsible for HIV-1 PICs to pass through the nuclear envelope in non-dividing cells, instead it requires concerted action of viral elements and host factors to direct PIC to actively pass through NPCs.

HIV-1 IN harbours several putative NLSs. Those NLSs including 156KELLKK, 186KRK, 211KELQKQITK, 235WKGPAKLLWK and 262RRKAK are rich in basic amino acids Lys and Arg and have been implicated in the nuclear import of HIV-1 PIC in various reports [16, 105, 109-111]. The bi-partite NLS (186KRK and 211KELQKQITK) and an overlapping bi-partite NLS (211KELQKQITK and 262RRKAK) have been proposed as functional NLSs which are required for Impα interaction and HIV-1 cDNA nuclear import [16, 110]. However, another study investigated IN two mutants including K186Q, Q214L/Q216L, which showed impaired nuclear accumulation of IN in transfected cells while retaining normal nuclear import of PIC [91, 111, 112]. Thus, the fate of the PICs does not correlate with the localization of IN.

HIV-1 nuclear import has involved two main routes to gain access into cellular nucleus including using Impα/ Importin β (Impβ) as adaptor/receptor and direct interaction with nuclearporins (Nups) [113]. Biochemical studies and genome-wide small interfering RNAs (siRNA) screening, yeast two-hybrid system have identified several nuclear import

receptors such as Importin 7 (Imp7), Impα3, Tranportin 3 (TNPO3) and Nups, proteins of the nuclear pore complex such as Nup62, Nup153 interacting with HIV-1 IN and assisting in its nuclear import and integration [109, 110, 114-116]. Our group had identified two regions 236KGPAKLLWK and 262RRKAK within CTD of IN required for binding Imp7 which is a member of the Impβ family, and mutations within these two regions impaired HIV-1 cDNA nuclear import [109]. Another Impβ family member TNPO3 was shown to physically interact with HIV-1 IN and participate in the nuclear import step [114]. However, the most recent studies reported that TNPO3 affects HIV-1 infection in the step after nuclear import but before integration through removal of remaining CA proteins and preparing PIC more readily integration into the host genome [117, 118]. Although the contribution of IN/Imp7 and IN/TNPO3 to HIV-1 nuclear import still remains unclear and controversial [114, 117-121], IN has also been shown to interact with different Impα subtypes and contribute to viral replication [16, 86, 110]. Among the four Impα subtypes (Impα1, 3, 5 and 7) Ao and colleagues tested, Impα3 knockdown exerts most potent effect on HIV-1 replication [86] and IN mutants defective for Impα3 binding showed reduced nuclear import of viral DNA and replication [110]. In line with this, the study done by Levin et al. showed that the interactions between IN with both Impα and TNPO3 are required for nuclear localization of IN and nuclear import of PIC [40].

The other pathway for IN-mediated nuclear import of PIC lies in the direct interaction of IN with Nups. The initial study reported that the direct interaction between Nup153 and IN mediated the nuclear import of HIV-1 [122], while a more recent study ascribed the

effect of Nup153 on nuclear import and integration steps of HIV-1 replication to uncoating of CA [123]. Nup62, another IN's interacting protein, is a nuclear pore protein and chromatin-binding protein [115]. However, Nup62 affects integration step rather than nuclear import of viral cDNA [115]. Some other Nups that function in the nuclear import of viral cDNA include Nup98, Nup358/ RAN binding protein 2, also known as RANBP2 [116, 124]. Nonetheless, whether nuclear import mediated by these Nups is through interaction with IN and how these nuclear pore structure proteins assist in the nuclear import and or integration of PIC still await further characterization.

In summary, despite the many cellular and viral determinants that have been proposed, the mechanism of HIV-1 nuclear import is not yet fully understood. Transport of a large complex like the PIC may require multiple contacts between PIC-associated factors with nuclear import factors as well as NPC constituents. From the information discussed above, it is conceivable that assisted by several cellular nuclear import factors and components from NPC, HIV-1 IN plays an important role in mediating nuclear import of PIC.

1.3.2.3 Integration

Integration proceeds in three steps: 3' processing, strand transfer and gap repair, occurring within a large nucleoprotein complex, referred to as PIC [81, 125]. During 3′ proessing which takes place in the cytoplasm of infected cells, IN removes a dinucleotide from each 3′ end of reverse transcribed viral DNA, generating overhanging 3′-hydroxyls attached to CA dinucleotides [126-128]. During strand transfer which occurs at the site of integration in the nucleus, the 3' ends of IN attacks the two strands of host chromosomal

DNA and the 3' processed viral DNA ends are covalently joined to the target DNA [129]. While IN dimer is sufficient for 3' process, 5' strand transfer requires at least IN tetramer formation: the two IN dimers from each end of viral DNA join together to form a tetramer [125, 130, 131]. All three functional domains are indispensable to carry out the 3' processing and DNA strand transfer [132, 133]. After 3' processing, suicidal autointegration might occur in which the reactive CA ends attack sites within the viral DNA, resulting in non-productive DNA circles instead of chromosomal integration [134, 135]. In addition, IN is able to catalyze a third reaction called disintegration which is the reverse of the strand transfer reaction [136]. The last step of integration (namely gap repair) involves the removal of two unpaired bases at the 5' ends of the viral DNA, filling in five-base single-stranded gaps between the viral DNA and the host DNA and finally ligation [137, 138]. While purified IN alone is able to perform the first two reactions 3' processing and 5' strand transfer in the *in vitro* assay, the gap repair step is believed to be accomplished by the host DNA repair pathway [77, 135]. Of note, mutations affecting enzymatic activities of IN refer as Class I mutations; Class II mutants perturb other steps of the retroviral replication such as virion morphogenesis, reverse transcription, etc [139].

Extensive studies have been conducted to investigate whether HIV-1 has a specific favor towards integration site selection. By utilizing human genome-wide mapping approaches, it has been shown that HIV-1 or HIV-based vector viral DNA favours integration into active transcription units in primary cells and transformed cell lines [70, 140, 141], while strongly disfavours the centromeric heterochromatin which are organized more compact than euchromatin and associated with repressed HIV transcription [142, 143]. Such

tendency is largely driven by the cellular protein LEDGF/p75 which tethers the lentiviral PIC to specific sites on the host chromosomes [140, 144]. In support of this notion, replacing chromatin binding domains (CBDs) from other proteins into LEDGF/75 and deletion of Pro-Try-Try-Pro (PWWP) domain of LEDGF/p75 redirected HIV-1 integration targeting [144, 145]. Interestingly, different retroviruses have differential patterns of the integration target sites. For instance, M-MLV tends to integrate into the promoter regions of genes in the human genome, while avian sarcoma-leukosis virus (ASLV) had a weak preference for active genes and showed no favour towards transcription start regions [141]. In consistent with this observation, LEDGF/p75 is not essential for M-MLV infection [140].

Notably, two cellular nuclear import factors TNPO3 and RANBP2 implicated in the nuclear import of viral PIC also impact on HIV-1 gene targeting sties. Integration targeting by TNPO3 or RANBP2 associated with gene dense regions, an integration pattern distinct from LEDGF/p75 which primarily targeted HIV-1 into active transcription units to promote efficient expression of viral proteins [146]. Since both of TNPO3 and RANBP2 are required for nuclear import of PIC as shown in previous studies [114, 116], the model of coupled nuclear import and integration targeting was proposed to explain this observation [116, 146] (the detailed description of the model can be found in the review [116]). The coupling is believed to be mediated by Nups, Imps, and putative tethering factors, the phenomenon which could explain some Nups are involved in either nuclear import or integration step such as Nup98, Nup62 [115, 116, 124], while some Nups affect both steps such as Nup153, Nup358/RANBP2 [123, 146]. Additionally, the tethering factor LEDGF/p75 might also play a role in the coupling of

nuclear import and integration [116]. The versatile functions of LEDGF/p75 including nuclear import of IN, IN-to-chromatin tethering, directing viral DNA integration into transcriptionally active regions suggest LEDGF/p75 as a promising candidate involved in the coupled nuclear import and integration [147, 148]. The detailed description of LEDGF/p75 will be reviewed in the following section.

1.3.2.4 The late stage of viral replication

While IN mainly affects steps between reverse transcription and integration during the early stage of viral replication, it also impacts on the late stage including transcription, packaging and processing of the viral polyprotein. It has been shown that degradation of IN is required for the integration-transcription transition of the viral replication cycle [82]. HIV-1 pol gene encoding IN region is essential for the proteolytic processing of viral Pr55gag polyprotein and viral particle morphogenesis [149].

1.3.3. Cellular cofactors of IN

HIV-1 is an obligatory intracellular pathogen of which all the known steps of different viral replication steps have involved the intimate virus-cell interaction. Interacting with a wide range of cellular proteins including PIC-associated and non-PIC-associated cellular factors confers versatility to IN functions that support the viral replication cycle. As discussed above, the PIC that formed after reverse transcription in the host cellular cytoplasm is a functional nucleoprotein complex consisted of viral DNAs, viral proteins such as IN, MA, nucleocapsid, RT, Vpr and various cellular proteins (reviewed by Al-Mawsawi LQ et al. [67]). By means of various approaches including yeast two-hybrid

system, co-immunoprecipitation (Co-IP), affinity tagged HIV-1 viral proteins and *in vitro* reconstitution of salt-stripped PIC activity, numerous cellular proteins within PIC have been identified [150-156]. The development of novel technologies utilizing *in vitro* integration assay along with mass spectrometry enable large-scale screening of host proteins associated with PIC [157]. Such cellular cofactors of IN indentified include LEDGF/p75, Ku, Ini1, HMGA1, BAF, Hsp60, *Polycomb* group EED protein, etc. IN has been shown to interact with some of these PIC-associated cellular factors and their interactions assist in the viral integration step. For example, Ini1, a component of chromatin remodelling SWI/SNF complex, binds HIV-1 IN and stimulates its DNA-binding activity [153]. BAF was also identified as a cofactor for HIV-1 IN, and it is able to restore integration activity of salt-denatured HIV-1 PICs while preventing autointegration in M-MLV [156, 158]. For the full list of these cellular cofactors for IN and their various functions, please see reviews [67, 68]. Here, only two of the cofactors of IN including LEDGF/p75 and Ku will be reviewed in this section.

1.3.3.1 LEDGF/p75

LEDGF/p75 was named after first isolation from lens epithelial cells of a cataract patient and was proposed as a growth factor for lens epithelial cells [159, 160]. However, this is a misnomer since it is neither lens-specific nor a growth factor. Rather, it is ubiquitously expressed and functions as a nuclear transcriptional coactivator which plays a protective role in the regulation of gene expression and cellular stress response [161]. LEDGF/p75 is a 530-aa protein encoded by the PSIP1 gene, which also encodes a 333-aa splice variant LEDGF/p52 that shares N-terminal 325 residues with LEDGF/p75 (Figure 4A).

LEDGF/p75 belongs to the hepatoma-derived growth factor (HDGF)-related proteins (HRPs). There are seven known HRPs in mammalian: HDGF, HRP1, HRP2, HRP3, HRP4 and LEDGF/p52/p75, all of which share the conserved N-terminal PWWP domain [162-164]. LEDGF/p75 has multiple domains which equip itself for nuclear import, chromatin binding and lentiviral IN binding [165] (Figure 4A). Chromatin binding ability of LEDGF/p75 which is essential for its function during HIV-1 infection is mediated by N-terminal half of the protein (the PWWP domain, NLS, two AT-hooks) and Charged regions (CRs) 1–3 [166-168]. The C-terminal of LEDGF/p75 harbours IN binding domain (IBD) encompassing aa 347-429 [148, 165, 169]. The LEDGF/p52 which lacks the IBD does not interact with IN and shows no effect on viral replication [147].

LEDGF/p75 was first revealed as a cofactor for HIV-1 IN by mass spectrometric analysis of immunoprecipitated epitope-tagged IN about a decade ago by Cherepanov and colleagues [147]. However, the smaller isoform LEDGF/p52 did not bind HIV-1 IN *in vitro* and in live cells [147]. Also the interaction of LEDGF/p75 with IN is lentiviral-specific including HIV-1, HIV-2, Simian immunodeficiency virus (SIV) and feline immunodeficiency virus (FIV), but not human T-cell leukaemia virus (HTLV), M-MLV or ASLV [170]. The mutual binding interface of IN/LEDGF/p75 has been resolved by alanine-scanning mutagenesis, X-ray crystallographic analysis and Nuclear magnetic resonance (NMR) spectroscopy [169, 171-173]. The CCD of IN is necessary and sufficient for binding LEDGF/p75 and it has been mapped to the residues A128-W132 and the region of I161- K173A in the catalytic core domain of IN [154, 169, 171-173].

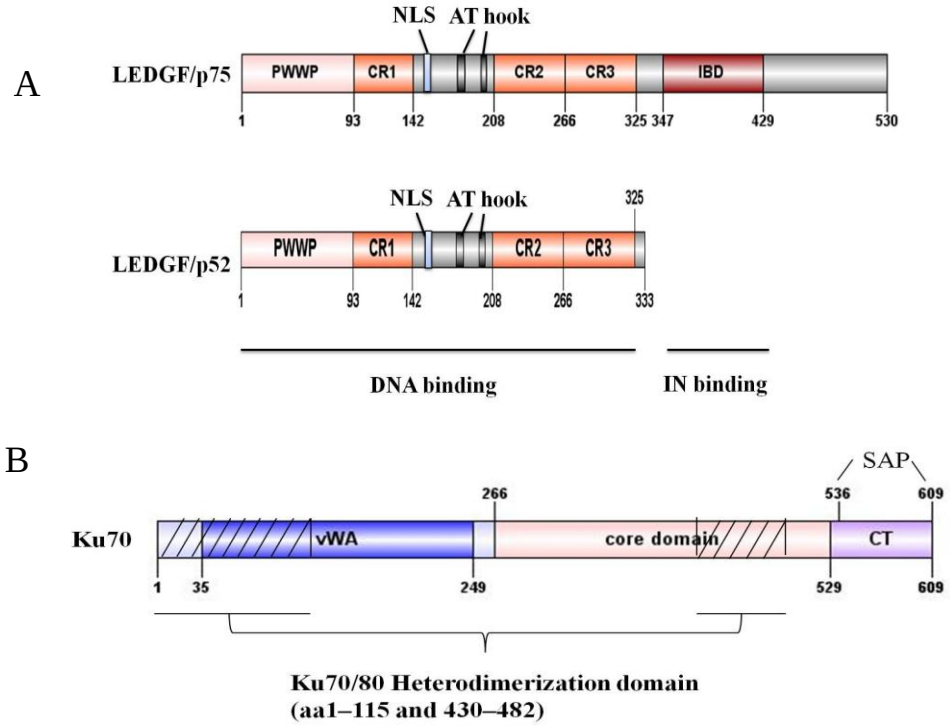

Figure 4 Domain organization of LEDGF/p75/p52 and Ku70. A) LEDGF/p75 contains an IBD in the C-terminus; and its N-terminal PWWP domain, the AT-hook domain, NLS and three relatively charged regions (CR1–CR3) are involved in chromatin binding. The smaller LEDGF/p52, produced by alternative RNA splicing, lacks the IBD and does not interact with IN. LEDGF, lens epithelium–derived growth factor; IBD, integrase-binding domain; PWWP, Pro-Try-Try-Pro; NLS: nuclear localization signal. B) Ku70 is consisted of the following regions: an N-terminal von Willibrand A domain (vWA), a central core domain and a C-terminal region (CT in the Figure). In Ku70, the CT region contains SAP (SAF-A/B, Acinus and PIAS) domain, which is involved in DNA binding.

As a functional component of PICs [170], the roles of LEDGF/p75 during lentiviral DNA integration have attracted increasing interest in recent years. Numerous studies have been performed to explore its functions and structures related to HIV-1 viral replication. The interaction of IN-LEDGF/p75 has several known impacts on IN and HIV replication. As discussed earlier, LEDGF/p75 serves as the IN-to-chromatin tethering factor driving PICs to transcriptionally active regions of host chromosomes [154, 155]. Recombinant full length or IBD of LEDGF/p75 was shown to enhance enzymatic activity of IN [147, 174]. LEDGF/p75 stabilized IN subunit-subunit interactions and promoted IN tetramerization [89]. In addition, the association between LEDGF/p75 and IN has been shown to protect IN from proteasomal degradation [175]. The LEDGF/p75 also plays a role in the nuclear transport of IN [147, 170]. Studies have revealed that lentiviral IN (HIV-1 and FIV) is redistributed from the nucleus to the cytoplasm in LEDGF/p75 knockdown cells but not for murine oncoretroviral IN [147, 170]. Moreover, LEDGF/p75 harbours a classical NLS 148GRKRKAEKQ156, and overexpression of a mutated form of this LEDGF/p75 NLS redirected HIV-1 IN in the cytoplasm [148, 165]. However, a subsequent study was not able to reproduce results. In this study, cellular localization of IN was not affected in LEDGF/p75 knockdown cells when treated with proteasome inhibitors [154]. The discrepancy can be explained by varied knockdown efficiency in the cells and residual endogenous LEDGF/p75 could still mediate nuclear translocation of IN. As a component of PIC, it was speculated that LEDGF/p75 might participate in the nuclear import of PIC. However, LEDGF/p75 does not seem to play a role in the nuclear localization of PIC as the depletion of LEDGF did not affect the ability of HIV or FIV to infect non-dividing cells while redistributing the nuclear localization of IN [170]. Nonetheless, the

involvement of LEDGF/p75 in lentiviral IN/PIC nuclear import remains ongoing areas of investigation.

While the importance of LEDGF/p75 in the stability and nuclear localization, chromatin association of IN has been well established, initial studies failed to observe significant reduction of HIV-1 replication in the LEDGF/p75 knockdown cells [167, 170]. It was reasoned that intracellular LEDGF/p75 level greatly exceeds the amount that are required to support HIV-1 integration [176]. This assumption was confirmed by the following studies using LEDGF/p75 knockdown and knockout techniques. The replication defect caused by LEDGF/p75 depletion occurred in the integration step while reverse transcription and nuclear import of PIC remained intact [120, 140, 177, 178]. Notably, HIV-1 integration in LEDGF/p75 knockout cells still remained 11% of the wild type level, suggesting that although LEDGF/p75 is important for IN-to-chromatin tethering and integration , it is not strictly essential [140]. In seeking the potential alternative factor for IN-to-chromatin tethering and nuclear targeting of IN, HRP-2 which is the only other human protein with a similar IBD domain was tested [165]. However, while HRP-2 was able to translocate IN from the cytoplasm to the nucleus in the LEDGF/p75 knockdown cells, it is not chromatin bound [165]. An elegant study was done by Schrijvers and coworkers demonstrated that HRP-2 knockdown in the human somatic LEDGF/p75 knockout cell line further reduced HIV-1 integration and replication [177]. Although this work provides convincing evidence that the residual replication in the absence of LEDGF/p75 is predominantly mediated by the same family member HRP-2, silencing both of LEDGF/p75 and HRP-2 still cannot completely abrogate HIV-1 integration [177]. Thus, it still remains unanswered whether HIV-1 could still integrate into host genome

without the cellular tethering cofactor or there is some other unidentified cellular partners involved in this process.

1.3.3.2 Ku

Ku70 is an evolutionarily conserved protein; it is found ubiquitously in eukaryotes and some prokaryotes such as Archaea and Bacteria [179-181]. As an abundant nuclear protein (0.5×10^5 molecule per nucleus) [182], Ku70 has been detected in different subcellular compartments such as nucleus [183, 184], membrane [185-187] and cytoplasm [188-190]. In the nucleus, Ku70 complexes with Ku80 and binds to the DNA ends at the break site as part of the Non-homologous end joining (NHEJ) pathway that repairs double-strand DNA breaks [191]; in the cytoplasm, cytosolic Ku70 binds and inhibits the pro-apoptotic protein Bax by preventing Bax translocating into the mitochondria [189, 192, 193]; Ku70 is also recruited to the cell membrane and functions in rickettsial entry into mammalian cells [185].

For most biological functions that Ku70 participates, Ku functions as a heterodimer consisting of Ku70 and Ku80, named according to their respective molecular weights of 70 and 80 kD. The NHEJ pathway begins with the recruitment of the Ku70/Ku80 heterodimer also known as Ku, followed by the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs) , Xrcc4 and DNA ligase IV [194]. Except for its best known roles in DNA repair and as a central component of the NHEJ pathway, Ku70 has also been implicated in many cellular processes including antigen-receptor gene rearrangement, mobile-genetic-element biology, V(D)J recombination of immunoglobulins, telomere maintenance, DNA replication, transcription, cell-cycle

control and apoptosis [180, 195]. Similar to LEDGF/p75, Ku70 is also a DNA binding protein in which it binds DNA specifically or non-specifically. As a DNA-repair protein, Ku70 can bind to any double-stranded DNA irrespective of sequence specificity or end configuration, including 5' overhangs, 3' overhangs, or blunt ends (see a review by Tuteja R, 2000) [195]. Meanwhile, Ku70 can also bind specific DNA sequences to affect gene transcription [196].

The full length of human Ku70 and Ku80 are 609 aa and 732 aa, respectively. Ku70 and Ku80 share high structural similarity although sequence alignment is only 14% identical [197]. Both of Ku70 and Ku80 have three structural domains composing of an N-terminal von Willibrand A domain (vWA), a central core domain and a C-terminal region (Structural organization of Ku70 is shown in Figure 4B) [180, 198]. vWA domain is protein–protein interacting domain. In Ku70, the C-terminal 5 kDa SAP (SAF-A/B, Acinus and PIAS) domain is involved in DNA binding or chromatin remodeling [194, 199, 200]. The heterodimerization domain in Ku70 was mapped to two regions aa 1–115 and 430–482 [201].

A recent study by Studamire et al. found that 12 cellular proteins, including Ku70, could bind to the INs of both the M-MLV and HIV-1 through screening with a yeast two-hybrid system [150]. Indeed, the involvement of Ku70 or Ku80 during HIV-1 replication has been documented in previous studies. As described in the three-step integration earlier, successful integration requires gap repair between viral DNA and host genome which is believed to be performed by host DNA-repair enzymes [202]. Two different host DNA-repair pathways have been suggested to fill in the gap during HIV-1 infection: the NHEJ and DNA damage-sensing pathways [203-205]. Studies have shown that the NHEJ

pathway is important for retroviral transduction or infection and for the cell survival of infected or transduced cells [204, 206-209]. For example, HIV-1-based vector transduction or infection was markedly reduced in cells deficient of Ku80, DNA-PKcs, Xrcc4 or ligase IV [206, 208]. Moreover, NHEJ activity is required for Two-long terminal repeat (2-LTR) circle formation, and Ku70 has been detected in M-MLV PIC [19, 208, 210, 211]. Ku80 was also shown to suppress HIV transcription by specifically binding to a negative regulatory element within the LTR [212]. All of these observations suggest that Ku70 or the Ku70/80 heterodimer may be involved in HIV-1 infection by affecting multiple steps of the viral replication cycle, such as integration.

1.3.4 Post-translational modification (PTM) of IN

As a single viral protein, HIV-1 IN interacts with numerous cellular cofactors in a temporally and spatially specific manner and exhibits multifunctional properties, which are tightly regulated by its different PTMs. To date, some of the known PTMs that modify IN that have been reported include: ubiquitination, SUMOylation, acetylation and phosphorylation. Among these modifications, cellular proteins either facilitate or counteract PTMs of IN such as p300 acetylates IN while LEDGF/p75 protects IN from ubiquitin-proteasomal pathway [114, 175]. PTMs on IN play various roles in the stability of IN, DNA binding, integration and infection of virus [114, 213-216]. This section will mainly focus on two of the PTMs of IN: ubiquitination and SUMOylation.

1.3.4.1 Ubiquitination

Among all these PTMs, ubiquitination is the first PTM discovered to modify IN in 2000 and perhaps the most well understood PTM [216]. The ubiquitin proteasome pathway

plays pivotal roles in many cellular processes through proteolysis of self or foreign proteins. Ubiquitination is a reversible PTM, in which ubiquitin (Ub) conjugates to the substrate proteins through covalent binding between the C-terminal Gly of Ub and the Lys residue of substrates. The process is accomplished through a cascade of enzyme reactions catalyzed by Ub activating enzyme (E1), Ub conjugating enzyme (E2) and Ub ligase (E3). Protein substrates can be mono-ubiquitinated or poly-ubiquitinated. The polyubiquitnation chain is formed through the isopeptide bond between C-terminal Gly of Ub and the internal Lys residue(s) of Ub which has been previously anchored to substrates. Five Lys residues within Ub including Lys 6, Lys 11, Lys 29, Lys 48, and Lys 63 among all seven internal Lys residues in Ub have been implicated in polyubiquitnation chain assembly [217, 218]. Lys-48-linked polyubiquitnation chain formation in the target protein is essential for the degradation of the protein, while Lys-63-linked polyubiquitination chain plays a role in the DNA repair [218, 219]. In the last step of Ub proteasome pathway, 26S proteasome recognizes and deubiquitinates the polyubiquitinated chain from the substrates and then protein substrates are degraded into small peptides. The Ub monomers are released from the polyubiquitination chain and recycled.

IN is known to be a metabolically unstable protein. In the absence of other viral proteins, IN in its natural status was undetectable in the tranfected cells, but gained expression with proteasome inhibitor treatment [216]. A few lines of evidences suggested that IN is subject to ubiquitin proteasome degradation pathway. After proteolytic cleavage from Gag-Pol precursor, HIV-1 IN bears a natural N-terminal Phe, the N-degron which accounts for its instability and replacing Phe with stabilizing residues (Met, Val, and Gly)

is able to stabilize IN [216]. A similar observation was obtained in another independent study [220]. Thus, degradation signal for IN fits into the general N-end rule pathway which has been found in mammals, plants and bacteria [221]. By using mutational analysis, Mousnier et al. reported that replacing Lys 211, 215, 219, or 273 with Arg in IN slowed down its degradation, suggesting that these Lys might be at least part of Ub targets [82]. In the same report, when all three Lys 211, 215, and 219 of IN were mutated into Arg in the virus, 60% reduction of HIV infectivity were observed [82]. However, further studies are still needed to verify whether the reduction of viral infectivity associates with prolonged degradation of these IN mutations in the context of viral infection.

There are two E3 ligase complexes which have been identified as E3 ligases for IN, and both of them contain RING fingers domain. Studies have indicated that proteins subject to N-end rule pathway are recognized by a set of ubiquitin ligase N-recognin (UBR) box-containing E3 ligases, which share ~70-residue zinc-finger-like domain termed UBR box [220]. And these E3 ligases recognize substrates through direct binding to N-terminal residue of protein substrates [222]. By using knockout and knockdown techniques, an early study suggested that mammalian UBR1, UBR2, and UBR4, as part of cellular E3 ligase complex, account for recognition and degradation of IN by the host Ub proteasome system (UPS) [220]. This model supports the notion that degradation of IN is mediated by N-degron-dependent proteolysis. A more recent study done by Mousnier et al. presented a distinct IN degradation model in which the prefoldin chaperone subunit von Hippel–Lindau binding protein 1 (VBP1) and the Cul2/von Hippel-Lindau protein (VHL) ligase interact with IN and mediate polyubiquitylation and proteasome degradation of IN

[82]. IN degradation was shown to take place after integration but before transcription during the viral life cycle [82]. Cul2/VHL Ub ligase complex, a multimeric ligase complex, is consisted of VHL, elongin C, elongin B, cullin 2 and RING-box protein 1. Within this complex, VHL binds and recognizes substrates whereas RING-box protein 1 confers ligase activity [223-225]. It still remains unknown how these two E3 ligases cooperatively or independently regulate IN proteasomal degradation pathway.

Although expression of authentic IN in the absence of other viral protein has revealed that HIV-1 IN is subject to rapid degradation, it is clear that IN is protected during a large portions of the viral life cycle. To date, two cellular proteins have been shown to interact with IN and impact on its half-life. Both of LEDGF/p75 and Human Rad18 (hRad18) have been revealed to play protective roles in the stability of IN [175, 226].

LEDGF/p75 was the first cellular protein reported to positively regulate IN expression [175]. LEDGF/p75 was discovered as IN-interacting protein by chemical cross-linking and mass spectrometry approach in 2003 [147]. Since then, multiple roles of LEDGF/p75 in the HIV-1 integration and infection have been revealed (see the content 1.3.3.1 above). The protective role of LEDGF/p75 has been supported by a few lines of evidences. For example, IN is less stable in LEDGF knockdown cells; artificial overexpression of LEDGF/p75 but not LEDGF/p52 increases IN expression [175]. It has been highlighted that the protection of LEDGF/p75 on IN from proteasome degradation relies on its interaction but irrespective of their nuclear or cytoplasmic localization [175]. A subsequent study undertaken by Mousnier A and coworkers suggested that LEDGF/p75 might protect IN from proteasomal degradation by masking its interaction site with VBP1 which is also an IN-interacting protein bridging IN with Ub E3 ligase Cul2/VHL [82]. In

spite of this, how LEDGF/p75 interacts and counteracts with the host UPS to protect IN remains largely unknown.

In addition, the DNA repair proteins hRad18 has been shown to interact with HIV-1 IN and stabilize IN. hRad18 along with ubiquitin conjugating enzyme E2 protein Rad6 are components of DNA post-replication repair pathway, involved in the repair of DNA lesions induced by mutagens such as UV light and chemicals during DNA replication [227, 228]. hRAD18 contains a RING finger domain shared by E3 ubiquitin ligases and it has been shown to target proliferating cell nuclear antigen (PCNA) for monoubiquitylation [227, 229, 230]. However, hRad18 has adverse effect on HIV-1 IN. The earlier study done by Mulder et al. has utilized the overexpression system and showed that hRad18 stabilized IN independent of its N-end rule [226]. So far, the mechanisms by which hRad18 confers protection on IN from proteasomal degradation is still unknown.

1.3.4.2 SUMOylation

In addition to ubiquitination, IN also undergoes small ubiquitin-like modifier (SUMO) modification, and this modification functions in the viral step between reverse transcription and integration during viral replication [214]. PTMs by Ub and a number of ubiquitin-like proteins (Ubls) such as Nedd8, SUMOs (SUMO1, 2 and 3), ISG15, FAT10, FUBI, UBL5, URM1, ATG8, and ATG12 are all reversible processes (See review [231]). Ub and Ubls covalently bind to protein substrates via an isopeptide bond, the linkage formed between C-terminal Gly in these protein modifiers and the ε-amino group of Lys in the substrates [232]. Similar to ubiquitination cascade, SUMOylation is catalyzed by

SUMO activating enzyme E1 (a heterodimer of Aos1 and Uba2), the unique E2 conjugating enzyme Ubc9 and a number of different E3 ligases such as PIAS and RanBP2 depending on the different substrates [233]. This whole process is reversed by specific isopeptidases, named sentrin-specific proteases (SENP). Mimicking ubiquitination, the substrate protein can be poly-SUMOylated or mono-SUMOylated, at single or multiple Lys targets. Unlike ubiquitination which primarily target subject proteins to proteasomal degradation pathway, outcomes for SUMO modification vary greatly from protein stability, cytosolic-nuclear translocation, antagonizing other post-translational modifications to transcriptional regulation [234]. SUMO proteins are ~10 kD in size and there are four subtypes (SUMO 1-4) in mammals all conserved among all the eukaryotic cells [234]. SUMO 1, 2 and 3 are ubiquitous in the cells and share globular ubiquitin-like shape while SUMO4 is only expressed in certain tissues and organs [234, 235]. While SUMO2 and SUMO3 are 96% identical to each other, SUMO1 only shares 45% amino acid identity with SUMO2/3. In the mammalian cells, most of SUMO1 exists as conjugated forms, while SUMO2 and SUMO3 are more likely to be free, non-conjugated monomers which are ready for quick conjugation and deconjugation processes [236].

SUMOs are covalently conjugated to protein substrates through the canonical four-amino-acid SUMO conjugation sites ψ-K-x-D/E (where ψ is a hydrophobic amino acid V/I/L/M/F and x is any amino acid). Of note, SUMO also binds substrates non-covalently through SUMO-interacting motif or SIM. SUMO conjugation or SUMOylation is generally used to refer to SUMO covalent binding while SUMO binding represents

SUMO non-covalent binding (See Figure 5). Conjugated SUMO proteins can provide a platform to recruit SIM-containing binding cofactors. Notable examples include: SIMs of human TRIM5α binding to SUMO-conjugated CA protein which restricts M-MLV infection [237], and RanBP2 SIM mediates its binding with the complex of RanGAP1/SUMO1 and Ubc9 [238]. Although the surrounding amino acids might also contribute to binding affinity or specificity to SUMO proteins, the core of SIM is consisted of four amino acids and three of them are hydrophobic residues V, I or L. The most common SIM is characterized as V/I-x-V/I-V/I or V/I-V/I-x-V/I/L where x can be any amino acid, in a parallel or anti-parallel orientation [239-241].

SUMOylation of IN is newly identified and the function is as yet poorly understood. Recently, Zamborlini et al. detected the SUMOyaltion of HIV-1 IN in both of *in vitro* and *in vivo* studies and showed that IN harbors three Ψ-K-x-D/E SUMOylation sites [214]. When these three critical Lys residues K46, K136 and K244 within the SUMO conjugation sites were mutated into similar positive charge Arg, SUMO modification of IN was drastically reduced, suggesting that these three amino acids are major SUMOylation sites [214]. However, SUMOylation of IN is not completely abolished with all the mutations, highlighting the presence of other SUMOylation sites. SUMOs can also bind to the Lys residue through inverted SUMOylation consensus motif E/DxKψ [232]. Indeed, there are two putative reverse (E/D)-x-K-Ψ SUMOylation sites in Lys71 and Lys 258 of IN. Thus, it would be interesting to address whether these inverted SUMOylation consensus motifs are bona fide SUMO conjugation sites. Meanwhile, we cannot exclude the possibility that SUMOylation of IN might take places on the Lys

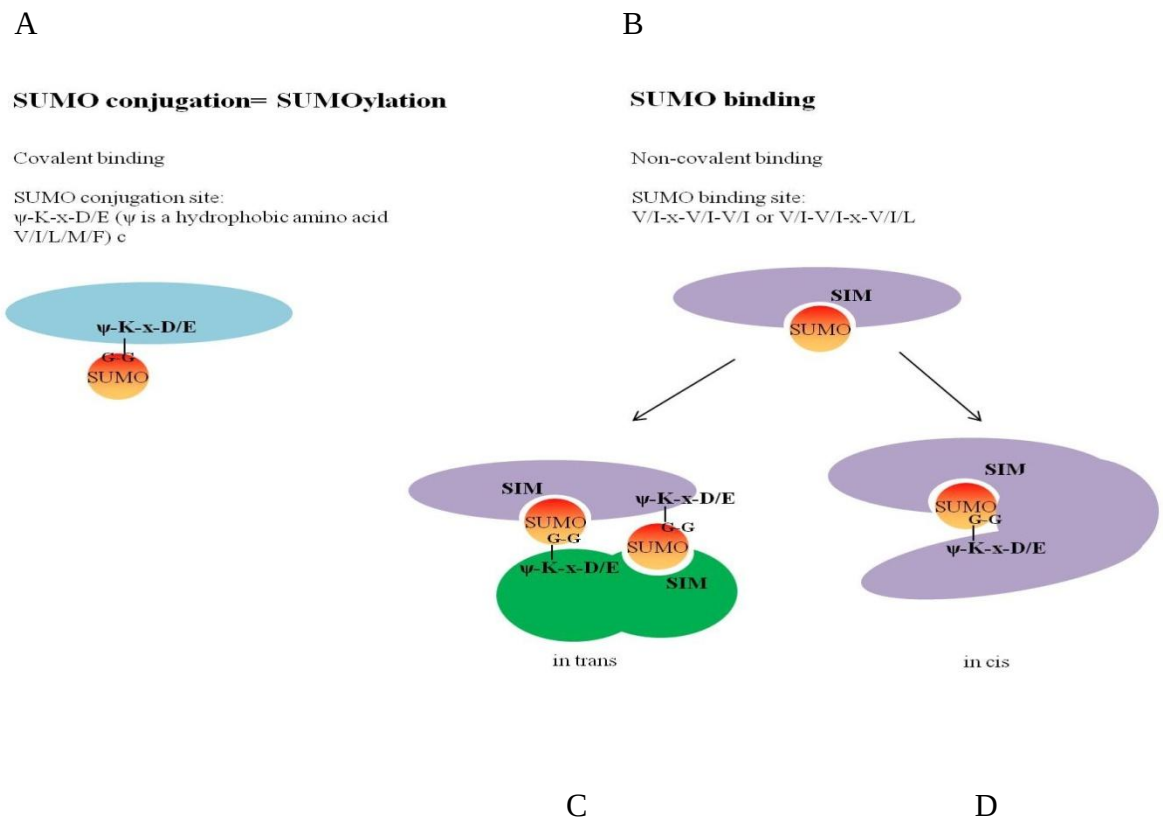

Figure 5 The SUMO conjugation and SUMO binding. A) SUMO conjugation or SUMOylation refers to covalent binding of SUMO moieties to protein substrates through SUMO conjugation sites ψ-K-x-D/E (left). During SUMOylation, an isopeptide bond is formed between C-terminal Gly of SUMOs and the ε-amino group of Lys in the substrates. B) SUMO binding represents SUMO non-covalent binding via SIM which bears four hydrophobic core residues of V/I-x-V/I-V/I or V/I-V/I-x-V/I/L. SUMO-SIM interaction can have various outcomes at the molecular level. SUMO-SIM interaction can be in trans (C) or in cis (D), thus recruiting cofactors (in trans) or affecting its own SUMOylation and inducing conformational change of protein substrates (in cis) (Reviewed in [242]). The SUMO-SIM interaction can serve as 'intra-molecular glue' to form multi-protein complexes (two proteins are colored purple and green respectively in C). SIM, SUMO-interacting motif; SUMO, small ubiquitin-like modifier.

residues within non-consensus SUMOylation region, as reported in other proteins such as PCNA and human E2-25K [243, 244]. Furthermore, SUMOs have three major isoforms (SUMO1-3), all of which have been shown to modify IN in the *in vitro* and *in vivo* studies. However, it is still unknown with respect to which SUMO subtype(s) preferentially target IN and whether IN is mono-SUMOylated or poly-SUMOylated.

This is the first clear evidence that IN is SUMOylated, but conclusive biochemical and functional data are still missing in terms of its impact on IN and HIV-1 viral replication. In seeking for the step(s) of which SUMOylation of IN takes place and the potential functions of this modification, the authors have ruled out the importance of SUMOylation of IN on nuclear translocation, protein stability, catalytic activity, LEDGF/p75 binding, multimerization and acetylation, ubiquitination of IN [214]. Although the exact functions of IN SUMOylation in HIV-1 infection still await further characterization, this report has underscored this new PTM of IN and defined the steps that SUMOyaltion of IN might take place between reverse transcription and integration in the course of the viral replication life cycle [214]. Despite limited data suggesting the relationship between HIV-1 IN and SUMO pathway, studies have found that SUMO E3 ligase RanBP2/Nup358 functions in the nuclear import of HIV PIC and integration site selection, and that SUMO2 is involved in the reverse transcription step of HIV-1, all of which are closely linked with the pleiotropic actions of IN [116, 146, 245]. At the molecular level, SUMO conjugation alters protein surface thereby promoting or disrupting protein-protein interaction (reviewed by Geiss-Friedlander et al.) [234]. Therefore, it is attractive to consider that SUMOylation of IN is involved in the interaction with its cofactors, due to the fact that other cofactors of IN such as

LEDGF/p75, Ku70, p300, Rad52 are also SUMOylated [246-249]. Studies are undergoing to understand the interplay between IN and components of the SUMO pathway [214].

1.3.4.3 Other PTMs of IN (acetylation, phosphorylation)

IN is acetylated by two cellular proteins p300 and GCN5 both *in vitro* and *in vivo* [250, 251]. Both of p300 and GCN5 are histone acetyltransferases (HATs), acetylating histone and non-histone proteins. While both of them acetylate IN at Lys residues K264, K266, and K273, K258 is exclusively modified by GCN5 [250, 251]. Consistent with this observation, deletion analysis on IN revealed that C-terminal last 24-44 aa of IN (aa 264-288 for p300 and aa 244-288 for GCN5) is the binding interface for both HATs. It is well known that IN CTD is essential for its DNA binding [84] and acetylation is able to increase DNA affinity through the neutralization of the positive charge of target Lys residues [252]. As expected, acetylation of IN by these two HATs increases IN binding to DNA in the *in vitro* UV-crosslinking DNA-binding analysis, and other functions of IN acetylation include positive regulation on IN enzymatic activity and integration within the context of viral replication [250, 251]. However, the importance of IN acetylation was questioned by another study in which HIV-1 virus containing the untagged IN mutant K(264/266/273)R is fully replication competent and the integration frequency is modestly impaired, even though they can reproduce and confirm the acetylation of IN [253]. The discrepancy was explained by the usage of epitope tag at the C-terminus of IN, which is close to the acetylation and mutation sites [253]. Therefore, it is still in debate as to the exact functions of IN acetylation especially during viral replication context.

Unlike ubiquitination, SUMOylation and acetylation which modify Lys residues, phosphorylation targets protein subjects on Ser, Thr or Tyr residues in eukaryotes [254]. Phosphorylation is also a reversible PTM, in which kinases phosphorylate protein substrates and phosphatases dephosphorylate proteins, acting as an on/off switch. Phospohrylation of IN was first appreciated by Manganaro L and coworkers in 2010 [213]. In this study, IN is shown to be phosphorylated at activated T lymphocytes by cellular kinase c-Jun N-terminal kinase (JNK) at Ser 57 [213]. IN mutant S57A which was abolished for its phosphorylation was more ubiquitinated, suggesting that phosphorylation and ubiquitination of IN antagonize one another in the regulation of its stability [213]. This raises the question as to how phosphorylation of IN impacts on its stability. To this end, the study discovered that the modified form of IN is recognized and stabilized by cellular peptidyl prolyl-isomerase enzyme Pin1, which is required for efficient HIV-1 integration and infection [213]. Pin1, which is a peptidyl-prolyl cis/trans isomerase, binds and catalyzes conformational changes of phosphorylated target proteins [213]. Indeed, experimental evidence revealed that Pin1 binding to phosphorylated IN lead to prolonged half-life of IN [213], this increased stability of IN may be ascribed to conformational change of IN catalyzed by Pin1. Such conformational change of IN might interfere with the accessibility of components from host ubiquitin-proteasome pathway such as ubiquitin, E3 ligase.

Like all viruses, the replication of HIV-1 must rely on host proteins due to its limited genome products. A number of proteins and/or cellular pathways are employed by HIV-1 to efficiently complete the replication cycle. Meanwhile, viral proteins such as HIV-1 IN and Tat are extensively modified by various PTMs, adding an additional level of

complexity to their functions during the replication cycles. These modifications are tightly controlled by various cellular proteins as discussed above. Furthermore, those PTMs including phosphorylation, ubiquitination, SUMOylation and acetylation often mingle with each other, either antagonistically or cooperatively, providing another level of cellular control. A better understanding on these complexes of virus-host protein-protein interactions and its regulated PTMs will certainly lead to exciting findings and potentially uncover a new intervention target to treat HIV-1 infection.

1.3.5 Anti-HIV-1 IN as a new class of ARV target

Since the commencement of Highly Active Antiretroviral Therapy (HAART) in 1996, the development of ARVs has undergone huge advances and it has significantly prolonged the life of HIV-infected patients as well as delayed AIDS progression. Currently, there are six major classes of ARV drugs being used to treat HIV-infected patients: nucleotide reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse-transcriptase inhibitors (NNRTIs), the PR inhibitors, the IN inhibitor raltegravir (RAL), the fusion inhibitor enfuvirtide (T-20), and the chemokine receptor 5 antagonist maraviroc [255]. Despite their great potency, these drugs face many drawbacks such as high cost, drug toxicities, drug-drug interactions and the emergence of resistance. Especially, resistance occurs in all ARV treatments. Thus, there is an urgent need for the development of new ARV drugs and drug classes to ensure efficacy, safety and low resistance to treat HIV-1 infected patients.

1.3.5.1 Strand transfer inhibitor (STI)

Theoretically, all stages of the viral replication life cycle can be potentially a pharmacological intervention for antiviral therapy. As discussed above, IN catalyzes the integration of viral DNA into the host genome, the essential step in its life cycle. This integration process is necessary for the establishment of provirus and allows the virus to replicate its genome using the host machinery. Blocking actions of IN or integration keeps the HIV genetic material out of the host genome, further preventing the production of infectious progeny viruses. Meanwhile, IN is a highly conserved viral enzyme, unique to retroviruses and no cellular homolog exist in the host cells [256]. Furthermore, IN inhibitors have potent antiviral activity against viruses with resistance to NNRTIs, NRTIs and PR inhibitors [257]. Therefore, HIV-1 IN is a promising target for ARV therapy and IN inhibitors could constitute a new class of drugs as part of combination regimens to treat infection by HIV-1.

Despite continuing efforts being made to explore the possibility of developing IN inhibitors, it has been a long way to develop an effective anti-IN drug. The first IN inhibitor RAL was approved in 2007, in comparison with the first NRTI known as Zidovudine or AZT in 1987 and the first HIV PR inhibitor saquinavir in 1995 [258, 259]. The three-dimensional (3-D) structures of two other viral enzymes, RT and PR, have been resolved and significant progress has been made towards developing structure-based inhibitors selective for RT and PR [258, 259]. The structural basis of viral DNA integration which allows modeling of the HIV-1 integration catalyzed by IN will aid in the development of ARV drugs. However, the rational design for IN inhibitors is hindered by the lack of the clear structure of full-length HIV-1 IN and IN in complex with both viral and target DNA [80, 98, 99, 260]. Studies attempted to crystallize HIV-1

IN have been obstructed by its poor solubility and propensity to form large aggregates under reaction conditions [98, 261]. To date, the crystal structures of all three individual domains of IN include: CCD, CCD plus CTD, NTD plus CCD, NTD and CCD complex with IBD of LEDGF/p75 have been solved with mutations to allow the improvement of the solubility while preserving its activity [98, 261]. Finally, a recent study disclosed the crystal structure of full length IN from another retrovirus family human foamy virus in complex with a pair of viral DNA and remodeled the inhibitory mechanism of two IN STIs RAL and elvitegravir (EVG) [256]. With these findings and future efforts towards understanding HIV-1 IN structure, it will provide invaluable information on how to optimize current STIs, how to prevent the virus from developing resistance as well as the development of next-generation IN inhibitors.

IN catalyzes both 3' process and 5' strand transfer, but all the IN inhibitors either available for use in the clinic or in clinical trials all target strand transfer step during integration and are therefore defined as IN STIs. The only available STI approved by the Food and Drug Administration (FDA) and is in use at the clinics is RAL. RAL has demonstrated good tolerance, great potency, and excellent pharmacokinetics. Two other STIs in advanced clinical development— EVG (Gilead Sciences) and S/GSK1349572 (GlaxoSmithKline) also showed efficacy in treatment-experienced patients [256, 262]. However, patients treated with RAL and EVG easily develop resistance, even one or two amino-acid mutation within the virus render 150-fold reduced susceptibility to these STIs [256, 262], while S/GSK1349572 has less tendency to develop resistance [263]. By analyzing data from *in vitro* passage experiments in the presence of RAL, sequence analysis in individuals receiving RAL and clinical HIV-1 isolates treated with RAL,

numerous mutations sites have been identified, with Q148HRK, N155H and Y143RC being the most occurrence resistant mutants (reviewed by Blanco [263]). Therefore, there is an urgent need to develop second-generation IN inhibitors.

1.3.5.2 Protein-protein interactions (PPIs) inhibitors: promising anti-HIV IN agents

PPIs play vital roles in various biological processes and have become a research focus in the current molecular biology field. Dysregulated PPIs can cause human diseases, either through the loss of an interaction or through the gain of an unnecessary protein complex. In cancer research, two essential PPIs including oncoprotein MDM2 binding to p53 and the Bcl-2 family members Bcl-2 and Bcl-XL binding to Bax have been targeted and their relative inhibitors have entered clinical trials in the treatment of tumours (reviewed by Zinzalla [264]). Currently, peptides and peptidomimetics derived from protein-protein interfaces, as well as, small-molecule PPI modulators are two main classes of PPI modulators being investigated [264].

PPI has also become an attractive class of molecular targets for ARV therapy. For example, HIV-1 PR is active when in the dimeric state. Thus, inhibition of the dimerization formation could inhibit its enzymatic activity, thereby leading to blocked budding and virion maturation and resulting in non-infectious viruses. The peptides derived from the dimerization surface of PR was first shown to inhibit HIV-1 PR dimerization in 1991 [265]. Continuous studies have led to the identification of a number of more potent, small-molecule dimerization inhibitors of HIV-1 PR [266, 267]. PPI can also be applied to HIV-1 IN and its catalyzed integration as integration is assisted by a

number of cellular factors. Instead of inactivating enzymatic function of IN, the alternative approach to block the HIV integration is to interfere with the key interaction of IN with its cofactors, consequently blocking integration and viral replication. HIV-1 IN has been shown to interact with dozens of cellular proteins, amongst which IN-LEDGF/p75 interaction is the most notable example and presents us a promising target for drug discovery. Disrupt of IN-LEDGF/p75 interaction aiming to block HIV-1 replication was rationalized based on previous published data on the IN-LEDGF/p75 interaction in the *in vitro* binding assay and *in vivo* cell-based assay, as well as infection block in LEDGF/p75 knockdown or knockout cells, etc [172, 178, 268-270].

There have been extensive efforts undertaken to facilitate drug development targeting IN-LEDGF/p75 interaction. For example, overexpression of the IBD of LEDGF/p75 could compete with endogenous LEDGF/p75 for binding to IN, leading to a potent defect in HIV-1 replication [178]. Peptides derived from the LEDGF/p75 loops (LEDGF/p75 aa353–378, LEDGF/p75 aa361–370, and LEDGF/p75 aa402–411) were found to bind IN and inhibit its catalytic activity and HIV-1 replication in infected cells [271]. The continuing study conducted more comprehensive investigation and confirmed the inhibitory effect of LEDGF/p75 aa361–370 on HIV infection *in vivo* using the mouse model [272]. The findings from these studies provided proof-of-concept that the IN-LEDGF/p75 interaction is a novel anti-HIV target. Structure-based and computer-aided drug research has been carried out and successfully identified 4-[1-(4-fluorobenzyl)-4-hydroxy-1H-indol-3-yl]-2-hydroxy-4-oxobut-2-enoic acid (CHI-1043) as dual inhibitor against IN strand transfer and IN-LEDGF/p75 interaction [273]. D77, a benzoic acid derivative, was found to bind CCD of IN, blocking the interaction between IN and the

LEDGF/p75 IBD and potently inhibiting HIV-1 replication [59]. By means of computational screening of a library consisting of 200,000 commercially available compounds and optimization of selected compounds, several 2-(quinolin-3-yl)acetic acid derivatives (LEDGINs) fit into the LEDGF/p75-binding pocket were identified, and they potently inhibited the IN-LEDGF/p75 interaction *in vitro* as well as HIV-1 replication in the infected cells [274]. Subsequent studies revealed that LEDGIN not only inhibited IN-LEDGF/p75, but also impaired integration activity by inducing premature multimerization of IN [192]. Notably, LEDGIN resistant virus harbour IN A128T mutation, which lies at the entrance of the LEDGF/p75-binding pocket, still retained sensitivity to anti-HIV drugs IN STI and AZT [274]. Thus, low cross-resistance between LEDGIN with other anti-HIV drugs suggests that targeting IN-LEDGF/p75 can complement current ARV regimens.

Another host protein Ini1, also known as hSNF5, was revealed as a binding partner for IN in a yeast two-hybrid screen [153]. It has also been explored as a potential therapeutic target. The researchers overexpressed a minimal IN-binding region of Ini1 (S6; comprising aa183−294) and observed potent inhibition of HIV-1 replication in a transdominant manner [275].

Furthermore, IN subunit-subunit interaction or multimerization has also been investigated as the potential target of novel anti-IN agents. A series of peptides derived from the dimerization interface of IN have been tested. Six peptides from the dimeric interface within CCD of IN were designed and synthesized, with three of them (95QETAYFLLKLAGRWP; 171HLKTAVQMAVFIHNFKR; 196AGERIVDIIATDIQ) inhibiting enzymatic activity and dimerization of IN [276]. Another peptide

167DQAEHLKTAVQMAVFIHNYKA187 encompassing the α5 helix and part of the loop between the α4 and α5 helices of IN was synthesized and showed potent inhibition on both 3' processing and 5' strand transfer, as well as oligomerization [277].

Taken together, currently both of the peptides either from virus or host origin and small-molecule inhibitors are being explored as novel anti-IN agents. Specifically, the host protein-origin peptide such as S6 fragment from Ini1 should not have cytotoxity and immunogenicity due to its host origin [275]. The development of peptides or small molecules that bind IN and inhibit its PPIs will enrich our understanding of the IN mode of actions, and lead to the development of new generation anti-IN agents.

Chapter 2 Hypothesis and Objectives

HIV-1 relies on the host cellular machinery to complete its replication cycle. Although the virally encoded enzyme IN alone is sufficient for the integration reaction *in vitro*, cellular cofactors are required in the regulation of multiple functions of IN and in assisting the integration step in the context of viral replication. Elucidation of the involvement of host factors in retroviral integration will contribute to the development of novel anti-HIV strategy based on inhibiting key interactions between IN and its cofactors.

The central hypothesis of this thesis is that HIV-1 IN utilizes various cellular cofactors to contribute to different aspects of IN and its related viral replication steps. The result of this research work will provide proof-of-concept for the feasibility of blocking HIV-1 replication though inhibition of the key interactions between IN and its cofactors.

To test this hypothesis, the four main objectives are listed below.

1) To investigate the mechanisms of IN-induced lethal phenotype in yeast by testing chromatin binding ability and cofactor binding affinity of lethal phenotype-defective IN mutants in **chapter 4**;

2) To characterize chromatin- and LEDGF/p75-binding abilities of IN by mutagenic analysis and investigate LEDGF/p75-independent chromatin binding of IN in **chapter 5**;

3) To evaluate the regulation of Ku70 in the stability or ubiquitination of IN and investigate its functional roles during HIV-1 infection in **chapter 6**;

4) To examine the importance of three putative non-covalent SUMO binding sites of IN in the regulation of SUMO conjugation of IN, cofactors binding and subcellular localization in **chapter 7**.

The first two objectives addressed the questions of how HIV-1 IN induces lethal phenotype in yeast cells and whether IN could mediate chromatin binding in LEDGF/p75-independent manner. The latter two aimed to understand how two PTMs (i.e. ubiquitination and SUMOylation) are involved in the interplay between HIV-1 IN with host cellular proteins. These four objectives are discussed in detail in chapter 4-7, respectively.

Chapter 3 Material and Methods

3.1 General Reagent

3.1.1 Cell lines and transfections

Human embryonic kidney 293T, the African green monkey kidney COS-7 cell lines and HeLa cell lines were cultured in Dulbecco's Modified Eagles Medium (DMEM) supplemented with 10% fetal calf serum (FCS) and 1% penicillin and streptomycin. Human CD4+ C8166 T cell line was maintained in RPMI-1640 medium supplemented with 10% FCS and 1% penicillin and streptomycin. 293T cells, COS-7 cells and HeLa cells were transfected with the standard calcium phosphate precipitation technique, as described previously [278, 279].

3.1.2 Yeast strains, culture media, and growth conditions

The protease-deficient *Saccharomyces cerevisiae* (*S. cerevisiae)* HP16 strain (*MATa ura3-52; his3Δ1; leu2; trp1Δ63; prb1-1122; pep4-3 prc1-407*) has been previous described [280]. Plasmid transformation was performed using the lithium acetate method [281]. The following culture media were used: 1) yeast complete medium YPD (1% yeast extract, 2% bactopeptone, 2% glucose) and 2) yeast liquid selective media: YNB lacking tryptophan (0.67% yeast nitrogen base without amino acids, 2% galactose or raffinose). Amino acids and bases (20-30 mg/1) were added as required. Solid selective media were obtained by supplementing liquid media with 2% bacto-agar. Yeast cells were grown at 30°C.

3.1.3 Plasmids

3.1.3.1 Expression plasmids for HIV IN and viral proteins:

To test the ability of the wild type HIV-1 IN to induce the lethal phenotype in the HP16 yeast strain, one HIV-1 IN yeast expression plasmid (p424Gal1-IN) was constructed by inserting a PCR-generated *Bam*HI-*Pst*I fragment containing the IN sequence into the high copy yeast expression plasmid p424Gal1. This plasmid contains a galactose-inducible Gal1 promoter and a tryptophan (Trp) selection marker. The different plasmids expressing mutant IN were generated using a two-step mutagenic PCR-based method [278] with primers containing the desired mutations. The amplified IN cDNA containing the specific mutations was then cloned into the p424Gal1 vector. All IN mutants were sequenced to confirm the presence of mutations.

To test the association between HIV-1 IN and cellular chromatin, different hemagglutinin (HA)-tagged IN expression plasmids (SVCMV-HA-IN) were constructed by fusing the IN cDNA to the 3' end of cDNA encoding the HA sequence (MASYPYDVPDYASL).

For the intracellular localization experiments and Co-IP assay, SVCMV-IN-YFP and SVCMV-YFP-IN, MA-YFP were constructed by using the same strategy described previously [109, 282]**.** SVCMV-YFP-IN mutants including EH170,1AA, EK170,3AA, HL171,2AA and HK171,3AA were constructed by PCR-based site-directed mutagenesis. The nucleotide sequences of the sense mutagenic oligonucleotides are as follows:

EH170,1AA, sense: 5'-AGATCAGGCTGCTGCTCTTAAGAC-3';
EK170,3AA, sense: 5'-GATCAGGCTGCACATCTTGCGACAGCAGT-3';
HL171,2AA, sense: 5'-AGGCTGAAGCTGCTAAGACAGC-3';
HK171,3AA, sense: 5'-AGGCTGAAGCTCTTGCGACAGCAGTAC-3'.

The amplified HIV-1 IN fragment was cloned in-frame at the 3' end of the EYFP cDNA

in a pEYFP-C1 vector (Clontech) at BglII and BamH1 sites. To construct pAcGFP-INwt/mut, each of the INwt/mut coding sequences was subcloned into pAcGFP1-C vector (Clontech) in-frame with the AcGFP coding sequence at BglII and BamH1 restriction sites.

To achieve high-level IN expression, a codon-optimized IN (INopt) cDNA was synthesized and cloned into the pUC57 vector (GenScript Co., Ltd.). To construct pAcGFP-INopt, the INopt fragment was excised from pUC57-INopt with BamHI and cloned in-frame at the 3' end of the pAcGFP1-C vector (Clontech) with the same restriction enzyme. To construct pAcGFP-INwt/mut, each of the INwt/mut coding sequences, including 1–230, 1–250, 1–270, 50–288, 112–288 and KR186,7AA, was amplified by PCR-based mutagenesis and subcloned into the pAcGFP1-C vector (Clontech) in frame with the Green Fluorescent Protein (GFP) coding sequence at the BglII and BamHI restriction sites [283].

To construct pAcGFP-INopt 3KR mutant (K46R/K136R/K244R) and 3VI mutant (V72A/I73A, I200A/V201A, I259A/I260A), the codon-optimized IN (INopt) 3KR and 3VI cDNAs were synthesized and cloned into the pUC57 vector by GenScript Co., Ltd. Then the INopt fragment was excised from pUC57-INopt with BamHI and cloned in frame at the 3' end of the pAcGFP1-C vector (Clontech) with the same restriction enzyme. The constructs of GFP-INopt mutants M1 (V72A/I73A), M2 (I200A/V201A), M3 (I259A/I260A), M1+M2 (V72A/I73A/I200A/V201A), M1+M3 (V72A/I73A/I259A/I260A) and M2+M3 (I200A/V201A/I259A/I260A) were synthesized through a two-step based PCR method using GFP-INopt wt as template which was described previously and cloned into pAcGFP-C vector using BglII and EcoR1 sites

[215]. All the mutants were confirmed by sequencing. The following primers were used:

INopt-BglII-5': 5-TAAGATCTTCCTGGACGGCA-3;

INopt-EcoR1-3': 5-GCTGAATTCTCAGTCCTCGTCCT-3;

INopt72,3AA-5': 5-AGGGAAAGGCTGCACTAGTGGCAGTG-3;

INopt72,3AA-3': 5-CACTGCCACTAGTGCAGCCTTTCCCT-3;

INopt200,1AA-5': 5-AGGAGAGAGGGCAGCTGACATCATC-3;

INopt200,1AA-3': 5-GATGATGTCAGCTGCCCTCTCTCCT-3;

INopt259,60AA-5': 5-AGCGACATCAAT GCAGCTCCTAGGCGGAAGG-3;

INopt259,60AA-3': 5-CCTTCCGCCTAGGAGCTGCATTGATGTCGCT-3.

3.1.3.2 Provirus or virus-related plasmids

The RT/IN/Env gene-deleted provirus (NL4.3Luc/ΔBg/ΔRI) has been previously described [282]. To complement RT/IN and Env defects of NL4.3Luc/ΔBg/ΔRI, a vesicular stomatitis virus G glycoprotein (VSV-G) expressor and a SVCMV-Vpr-RT-IN fusion protein expressor [109, 282] were used in this study. To test the effects of different IN mutants on viral infection, cDNAs encoding the IN mutants, including V165A, A179P, KR186,7AA, D64E, EH170,1AA, EK170,3AA and HL171,2AA were introduced into the SVCMV-Vpr-RT-IN expressor by PCR-based method as described before [282].

The HIV-1 proviruses pNL4.3-GFP, HxBru or HxBru-IN-HA were described earlier [279]. CMV-Vpr-RT has two stop codons TAGTGA in replace of first six nucleotides in IN sequence of CMV-Vpr-RT-IN and the sequence was confirmed by sequencing. To construct CMV-Vpr-RT-IN-ProLabel (Vpr-RT-IN-PL) plasmid, two-step based PCR method was used. ProLabel tag sequence was amplified from pProLabel-C Vector from

ProLabel™ Detection Kit II (Clontech Laboratories, Inc.) and inserted after IN sequence in CMV-Vpr-RT-IN plasmid with IN stop codon and ProLabel start codon removed. The following primers were used:

RT NheI 5': 5-GCAGCTAGCAGGGAGACTAA-3;
R-RT-IN-ProLabel pstI 3': 5-GTCGACTGCAGAATTCGAAGCTTATTC-3;
IN-ProLabel 5': 5-AGACAGGATGAGGATAGCTCCAATTCACTG-3;
IN-ProLabel 3': 5-CAGTGAATTGGAGCTATCCTCATCCTGTCT-3.

3.1.3.3 Constructs of cellular proteins

To construct SVCMVin-T7-LEDGF, the LEDGF cDNA derived from a pFT-1-LEDGF plasmid [167] was cloned into a SVCMVin vector [284]. The pFT-1-LEDGF plasmid was kindly provided by Dr. A. Engelman through the AIDS Research Reference Reagent Program, Division of AIDS, NIAID, NIH. LEDGF/p75 was cloned into the pProLabel vector in-frame downstream of the ProLabel tag named pProLabel-LEDGF.

LEDGF/p75 SUMOylation defective mutant T7-LEDGF/p75 4K mutant (K75R/K250R/K254R/K364R) was constructed through mutagenesis from T7-LEDGF/p75 wt which was described earlier by two-step-based PCR method [215, 285]. BamH1 and HindIII were used to clone into SVCMVin-T7 vector. The following primes were used:

LED-BamH-5': 5' CCGGATCCACGACTCGCGATTT 3'
LED-HindIII-3': 5' CGGATCAAGCTTCGAATTCCT 3'
LED-K75R-5': 5' CCAAATAAAAGCCGCGGTTTTAATG 3'

LED-K75R-3’: 5' CATTAAAACCGCGGCTTTTATTTGG 3'

LED K250RK254R-5’: 5' GAGCCGGATAGAAAAGAGGGGAGGAAAGAAGT 3'

LED K250RK254R-3’: 5' ACTTCTTTCCTCCCCTCTTTTCTATCCGGCTC 3'

LED K364R-5’: 5' TTCACTCCGGATTGATA 3'

LED K364R-3’: 5' TATCAATCCGGAGTGAA 3'

The Ini1 expressor pCGN-HA-Ini1 used in this study was described previously [286].

Untagged human full-length Ku70 cDNA in the pCMV6-XL5 vector was purchased from OriGene Technologies Inc. To construct SVCMV-T7-Ku70, a Ku70 cDNA without a start codon was amplified and cloned into the SVCMVin-T7 vector at the BamHI and NotI restriction sites. The T7-Ku70 truncation mutants (1–263, 1–430, 226–609 and 430–609) were obtained using the same strategy. The nucleotide sequences of the mutagenic oligonucleotides are as follows:

5’Ku70-BamH1: 5-TAGCCGGATCCTCAGGGTGGGAGTCATATTA-3;

3’Ku70-Not1: 5-T ATATGCGGCCGCTCAGTCCTGGAAGTGCTT-3:

Ku70-263-Not1: 5-AGATGCGGCCGCCTAGAGCTTCAGCTTT-3;

Ku70-430-Not1: 5-TATGCGGCCGCTCATGGAGGAGTCACCTGAAT-3;

Ku70-226-BamH1: 5-TATGGATCCGATGAGGACCTCA-3;

Ku70-430-BamH1: 5-ATATGGATCCCCAGGCTTCCAGCT-3.

HA-tagged Ubiquitin (HA-Ub) and mutants HA-UbK48R, HA-UbK63R were provided by Dr. Gottlinger HG (UMass Medical School, Boston, USA) and described previously [287]. HA-SUMO3 (plasmid 17361) was obtained from Addgene [288]. HA-Ubc9 wt was constructed from pcDNA3-V5-Ubc9 wt digested with BamH1 and XhoI and cloned

into CMV-HA vector. pcDNA3-V5-Ubc9 was a generous gift from Dr. Ronald Hay (University of St. Andrews, St. Andrews, UK).

3.1.4 Antibodies and chemicals

Epitope tag antibodies: The rabbit anti-GFP (Molecular Probes), mouse anti-HA antibodies (Molecular Probes), the rabbit anti-HA antibody (Sigma), anti-T7 monoclonal antibodies (Novagen) were used either for fluorescence assay, immunoprecipitation (IP) or Western Blot (WB). Horseradish peroxidase (HRP)-conjugated anti-GFP antibody (Molecular Probes), HRP-conjugated anti-HA antibody (Miltenyi Biotec.), and HRP-conjugated anti-T7 antibody (Novagen) were used for WB detection.

HIV and IN antibodies: Anti-IN antibodies were kindly provided by Dr. Grandgenett through the AIDS Research Reference Reagent Program, Division of AIDS, NIAID, NIH. Anti-HIV p24 monoclonal antibody used in this study were previously described [284].

Antibodies recognizing cellular proteins: The mouse monoclonal antibody against yeast β-actin (ab8224) was purchased from Abcam and mouse anti-α-tubulin was form Sigma. And the monoclonal anti-Nup62/p62 antibody was purchased from Sigma. Rabbit anti-LEDGF/p75 (Bethyl Laboratories, Inc.), mouse anti-Ku70 (Abcam), mouse anti-Ku80 (Abcam) antibodies were used as primary antibodies in WB or IP assay.

Secondary antibodies: HRP-conjugated donkey anti-rabbit IgG and sheep anti-mouse IgG were purchased from Amersham Biosciences and GE healthcare, respectively. Fluorescein isothiocyanate (FITC)-conjugated anti-rabbit antibody (Kirkegaard & Perry Laboratories (KPL) was used in the fluorescence assay.

Chemicals: The WB-detection Enhanced chemiluminescence (ECL) kit was purchased from PerkinElmer Life Science (Boston, MA). 4'-6-Diamidino-2-phenylindole (DAPI) was purchased from Molecular Probes, Invitrogen. NP-40 was from Roche. Protease Inhibitor Cocktail Set III, MG-132 and puromycin were obtained from Calbiochem. Subtilisin was purchased from Sigma.

3.2 General Methods

3.2.1 Evaluation of the lethal phenotype induced by HIV-1 IN in the HP16 yeast strain

The experimental procedures to evaluate protein expression and yeast growth arrest activity were described previously [278]. Briefly, HP16 cells transformed with p424Gal1 or p424Gal1-IN wild type or mutant plasmids were first grown in an IN non-inducible selective medium (Trp^-, 2% raffinose (raf+)) for 2 days. Then, equal amounts of transformed HP16 yeast cells were inoculated into IN-non-inducible (Trp^-, 2% raf+) or IN-inducible (Trp^-, 2% galactose (gal+)) media for the liquid assay. After cultivation while shaking at 30°C for 24-36 h, yeast growth was monitored by measuring the culture density spectrophotometrically at 600 nm. Meanwhile, equal amounts of transformed yeast cells were serially diluted and spotted onto either IN non-inducible or inducible agar plate for the "drop test". After incubation for 3 to 5 days, yeast growth was recorded photographically. To detect IN expression in yeast, equal amounts of transformed yeast cells were grown in IN-inducible media for 6 h, then cells were pelleted by centrifugation and lysed in RIPA buffer by vortexing with glass beads for 1 min on ice, four times. Supernatants were collected and IN was immunoprecipitated with anti-HIV

antibodies. Immunoprecipitates were then resolved by 12.5% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) followed by WB using rabbit anti-IN antibody.

3.2.2 Yeast chromatin-binding assay

HP16 cells harboring either p424-Gal1-IN wild type or p424-Gal1-IN encoding the V165P, A179P or KR186,7AA mutants were grown at 30°C in IN-non-inducible (Trp^-, 2% raf+) media to a cell density of 0.25 A600. Cells were then induced with IN-inducible (Trp^-, 2% gal+) media for 3 h. Cells were spheroplasted and fractionated, as described previously [289] with minor modifications. Briefly, cells were resuspended in spheroplasting buffer (1% yeast extract, 2% peptone, 0.2% galactose, 50 mM KH_2PO_4/K_2HPO_4 (pH 7.5), 0.6M Sorbitol, 10 mM DTT) containing 20 µl of 10 mg/ml of Zymolyase 100T. Spheroplasts were centrifuged for 3 min at 1000 x g and then resuspended in 200 µl of ice-cold lysis buffer (0.1M Pipes-KOH, 10mM DTT, pH 9.4) containing proteases inhibitors and lysed in 1% Triton X-100, yielding whole cell extracts. Chromatin and non-chromatin fractions were separated by centrifugation for 15 min at 15 000 x g and subjected to WB analysis.

3.2.3 Chromatin binding assay in 293T cells

The association of HIV-1 IN with chromatin in mammalian cells was analyzed using a previously described chromatin binding assay [290]. Briefly, 293T cells were transfected with different SVCMV-HA-IN mutants. At 40 h post-transfection, cells were lysed for 15 min on ice with cold CSK I buffer (10 mM Pipes pH 6.8, 100 mM NaCl, 1 mM EDTA, 300 mM sucrose, 1 mM MgCl2, 1 mM DTT) supplemented with 0.5% (v/v) Triton X-

100 and protease inhibitors. Cell lysates were centrifuged at 500 x ***g***, for 3 min at 4°C to separate Triton-soluble (S1) and non-soluble (P1) fractions. Half of the S1 fraction was further lysed in RIPA buffer (150 mM Tris–HCl, pH 8.0, 150 mM NaCl, 0.5% DOC, 0.1% (w/v) sodium dodecyl sulfate (SDS), 1% (v/v) NP-40). The P1 fraction, which contains chromatin-bound, nuclear matrix-bound and insoluble proteins, was divided into two equal portions. One half was resuspended in RIPA buffer (the P1 fraction). The remaining half was resuspended in CSK II buffer (10 mM Pipes pH 6.8, 50 mM NaCl, 300 mM sucrose, 6 mM MgCl2, 1 mM DTT) and treated with DNAse (10 unit) for 30 min. It was then extracted with 250 mM $(NH_4)_2SO_4$ for 10 min at 25°C and centrifuged at 1200x***g*** for 6 min at 4°C. The supernatant (S2 fraction, containing DNase-released chromatin-associated proteins) and pellet (P2, containing insoluble, cytoskeletal, and nuclear matrix proteins) were collected and resuspended in RIPA buffer. All fractions were immunoprecipitated with anti-HA antibodies followed by a WB with the same antibody.

To simplify the assay, after transfection of YFP-INwt/mut plasmids only S1 (non-chromatin-bound) and S2 (chromatin-bound) fractions in chapter 5 were analyzed by immunoprecipitation using an anti-GFP antibody and detected by WB with the same antibody. Protein bands in each fraction were further quantified with the software Quantity One (Bio-Rad), and the values are expressed as a percentage of chromatin-bound YFP-IN to total input, which consists of YFP-IN present in both S1 and S2.

3.2.4 Co-immunoprecipitation (Co-IP) assay and chemiluminescent Co-IP assay

The interaction between IN and the cellular proteins Ini1 or LEDGF/p75, Ku70 was

verified by Co-IP. 293T cells were first cotransfected with different IN plasmids including SVCMV-IN-YFP or YFP-IN, GFP-IN with pCGN-HA-Ini1 or SVCMVin-T7-LEDGF or SVCMVin-T7-Ku70. At 48 h post-transfection, cells were lysed with 0.25% NP-40 prepared in RPMI-1640 medium and a protease inhibitor cocktail (Roche) on ice for 30 min and the extracts clarified by centrifugation at 14,000 rpm for 30 min. The cell lysates were precleared for at least 2 h at 4°C. To test the interaction between IN and Ini1, IN-bound Ini1 was detected by immunoprecipitation with anti-GFP antibody overnight and incubation with protein A sepharose for another 2h. After washing the beads for five times, the bound proteins were detected by a WB using an anti-HA antibody. To detect the IN/LEDGF complex, the extracts were subjected to immunoprecipitation with rabbit anti-GFP antibody and followed by a WB using a mouse anti-T7 antibody. The IN/Ku70 interaction was detected by IP with a rabbit anti-GFP antibody and immunoblotting with anti-Ku70 or anti-T7 antibodies in WB. Meanwhile, the unbound HA-Ini1, T7-LEDGF and T7-Ku70 proteins in the remaining cell lysates were checked by an immunprecipitation with anti-HA or anti-T7 antibodies followed by a WB with the same antibody. Alternatively, 10% of transfected cells were lysed in 0.5% NP-40, and the lysates were used to detect the expression of YFP-INwt/mut and T7-LEDGF/p75, T7-Ku70 by WB using anti-GFP and anti-LEDGF or anti-T7 antibodies, respectively. The expression of IN-YFP or YFP-IN was checked by a WB with anti-GFP-HRP antibody. To increase GFP-IN stability, 10 μM MG-132 was added 12 h prior to cell lysis for Co-IP.

The chemiluminescent Co-IP assay was performed according to manufacturer's instructions. After AcGFP1-INwt/mut or AcGFP1-C and ProLabel-LEDGF, or ProLabel-Ku70 fusion protein expression plasmids were co-transfected in 293T cells for 48 h, the

cells were collected and lysed in 0.25% NP-40 lysis buffer and co-immunoprecipitated with anti-GFP polyclonal antibody and protein A sepharose. After washing the beads five times with the lysis buffer, the immunoprecipitates were resuspended in lysis/complementation buffer and transferred to a well in a 96-well assay plate (Costar, Corning, NY). To each well, the substrate mix was added, and ProLabel activity was measured using the POLARstar OPTIMA multidetection microplate reader (BMG Labtech, Ortenberg, Germany).

To examine the IN/Ku70 interaction in HIV-1-infected cells, HIV-1 (HxBru or HxBru-IN-HA)-infected C8166 T cells were lysed with 0.25% NP-40 and immunoprecipitated with anti-HA antibody followed by WB with anti-Ku70 antibody to detect IN-bound endogenous Ku70.

3.2.5 Detection of ubiquitination of IN in the absence or presence of Ku70

To determine the ubiquitination level of HIV-1 IN in the absence and presence of Ku70, 293T cells were cotransfected with GFP-IN and HA-Ub wild type (wt) or mutants K48R and K63R with and without T7-Ku70wt or T7-Ku701-430. After 48 h, cells were lysed in 199 medium containing 0.25% NP-40 and a protease-inhibitor cocktail and immunoprecipitated with anti-GFP antibody. Then, the precipitated complexes were run on a 10% SDS-PAGE gel and analyzed for the presence of HA-Ub by WB with HRP-conjugated anti-HA antibody. Simultaneously, GFP-IN was detected by immunoblotting the same membrane with HRP-conjugated anti-GFP antibody. And protein band intensity was quantified using Quantity One 1-D Analysis Software (Bio-Rad Laboratories).

3.2.6 Immunofluorescence assay

COS-7 cells or HeLa cells were grown on glass cover slips (12 mm^2) in 24-well plates for 24 h and then transfected with different IN expression plasmids (CMV-YFP-IN or GFP-INopt). After 48 h, cells on the cover slip were fixed and permeabilized for 30 min in methanol/acetone (1:1 ratio) at room temperature. The glass cover slips were incubated with a primary rabbit anti-GFP antibody followed by a secondary FITC-conjugated anti-rabbit antibody. Nuclei were stained with DAPI. Cells were visualized on a Carl Zeiss microscope (Axiovert 200) with a 63x oil immersion objective.

To test the effect of Ku70 knockdown on the expression of IN, HeLa cells were first transfected with Ku70-specific siRNA oligonucleotides or nontargeting random siRNA (siNC) for 48 h and further transfected with GFP-INopt for another 48 h with or without MG-132 (10 μM) treatment. Then immunofluorescence assay was performed as described above.

3.2.7 Transient knockdown of LEDGF/p75 or Ku70

To transiently knockdown LEDGF/p75 and Ku70, RNA interference (RNAi) for LEDGF or Ku70 and scrambled RNAi were purchased from Invitrogen. 293T cells or HeLa cells were seeded in a 6-well plate for 24 h and then cells were transfected with siRNA oligonucleotides (20 nM for siLEDGF and 5nM for siKu70) (Stealth RNAi; Invitrogen) directed against PSIP1/LEDGF/p75 or Ku70 mRNA using Lipofectamine 2000 (Invitrogen) or Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen). Synthetic siRNA for LEDGF/p75 targeting nucleotides 541 to 565 of PSIP1HSS146003 (5'UAAUGAAGGUUUAUGGGAGAUAGAU3'). The sense primer for siKu70 is 5-

GAUCCAGGUUUGAUGCUCAtt-3, targeting Ku70 nucleotides 1,094 to 1,112. In parallel, a scramble siRNA was used as negative control (siNC). RNAi-mediated knockdown of endogenous LEDGF/p75 and Ku70 was monitored by WB at different time points using corresponding antibodies.

3.2.8 The production and transduction of lentivirus vector containing LEDGF/p75 and Ku70 short hairpin RNA (shRNA)

To produce stable LEDGF/p75 gene knockdown 293T cell lines, the pLKO.1 lentiviral vector comprising siRNA hairpin targeting nucleotides of LEDGF/p75 mRNA was purchased from Open Biosystems (Waltham, MA, USA). The hairpin structure contains a 21-bp stem, 5-nt loops, and 5' CCGG and 3' TTTTTG overhangs. The shRNA sequence RHS3979-97063117 targets the corresponding LEDGF/p75 mRNA nucleotides 860-880, and its stem-loop sequence was CCGGGCAGCTACAGAAGTCAAGATTCTCGAGAATCTTGACTTCTGTAGCTGC TTTTTG. shRNA pLKO.1 vector containing shRNA targeting the Ku70 mRNA: 5-CCGGCGACATAAGTCGAGGGACTTTCTCGAGAAAGTCCCTCGACTTATGTCGT TTTTG-3 (Oligo ID: TRCN0000039608) was also purchased from Open Biosystems. The lentiviral particles harboring LEDGF/p75 or Ku70 shRNA were produced by co-transfecting the shRNA pLKO.1 vector, packaging DNA plasmid Δ8.2 and VSV-G plasmid into 293T cells. After 48 h, supernatants containing lentiviral vectors were pelleted by ultracentrifugation (32,000 rpm at 4 °C for 1 h) and stored in aliquots at – 80 °C.

To obtain stable LEDGF or Ku70 shRNA expressing cell lines, 293T cells or C8166 T

cells were transduced with the shRNA LEDGF or shRNA Ku70 lentiviral vector for 48 h and then selected with 2 μg/mL puromycin for one week. Silencing of LEDGF/p75 or Ku70 was determined by WB analysis with an anti-LEDGF or anti-Ku70 antibody. Detection of endogenous beta-actin was used for normalization of sample loading. The pLKO.1 vector without the shRNA sequence (empty vector) was introduced into cells by the same method as a negative control.

3.2.9 *in vivo* SUMOylation Assay

Detection of IN SUMOylation using immunoprecipitation analysis was described previously [291], with minor modifications. 293T cells were cotransfected with HA-SUMO3 and AcGFP-INwt/mut for 40 h. The cells were harvested and washed in cold Phosphate buffered saline (PBS) once and then lysed in 150μl lysis buffer (0.15 M Tris-HCl, pH 6.7, 5% SDS, and 30% glycerol), which is then diluted 1:10 in PBS/0.5% NP40 plus complete protease inhibitor (Roche). Prior to cell lysis, pelleted cells were added 10 unit/ml Benzonase (Novagen, Billerica, MA, USA) in 2mM $MgCl_2$ for 20 min to reduce the cellular viscosity. Cell lysates were centrifuged at 16,000×g for 10 min at 4°C to remove cellular debris. The supernatants were first precleared at 4°C for 30 min by protein G-Sepharose. The cell lysates were immunoprecipitated with anti-GFP antibody for 2 h and followed by the protein A-Sepharose for another 2 h at 4°C. The beads were collected and washed with PBS/0.5% NP40 buffer five times. The bound proteins were eluted with 4x Laemmli buffer and separated on SDS-PAGE gel. SUMO conjugated IN was detected by anti-HA antibody in WB.

3.2.10 Subcellular protein fractionation

293T cells were transfected with AcGFP-INopt wt or 3VI in 6-well plate for 48 h. Cells were harvested and proteins were sequentially extracted using differential detergents, yielding cytoplamic, nuclear and chromatin-bound fractions using a Thermo Scientific Subcellular Protein Fractionation Kit according to manufacturer's protocol. Each fraction of proteins was subject to WB analysis using anti-GFP antibody.

3.2.11 Single cycle viral replication and infection

A VSV-G pseudotyped single-cycle replicating virus was produced in 293T cells as described previously [292]. Briefly, 293T cells were co-transfected with an RT/IN/Env-deleted HIV-1 provirus NLlucΔBglΔRI, each CMV-Vpr-RT-IN (wt/mutant) expression plasmid and a VSV-G expression plasmid. After 48 h post-transfection, viruses were collected and concentrated from the supernatants by ultracentrifugation at 35,000 rpm for 2 h. Virus titers were quantified using HIV-1 p24 Antigen Capture Assay Kit (purchased from the NCI-Frederick AIDS Vaccine Program).

To test the effect of the IN mutants on viral infection, equal amounts of VSV-G-pseudotyped, single cycle replicating viruses (adjusted by virion-associated p24 levels) were used to infect C8166 T cells overnight at 37 °C. At different time points after infection, 1 x 10^6 cells from each sample were collected and lysed with 50 µL of luciferase lysis buffer (Fisher Scientific Inc., Rockford, IL USA). A 10 µL aliquot of cell lysate was subjected to the luciferase assay by using a POLARstar OPTIMA (BMG LABTECH, Germany), and the luciferase activity was valued as relative light units (RLU).

3.2.12 ProLabel detection assay to check metabolism of IN under infection

condition

To test the effect of Ku70 on IN during HIV infection, VSV-G pseudo-typed HIV single cycle virus containing ProLabel tag fused to C-terminal of IN was generated to quantify IN expression under HIV infection. NL4.3lucΔBglΔRI was cotransfected with Vpr-RT-IN-ProLabel (Vpr-RT-IN-PL) and VSV-G expressor into 293T cells to generate VSV-G pseudotyped HIV-1 single cycle IN-PL virus. The viruses were used to infect shKu70-knockdown or empty-vector transduced C8166T cells for 3 h. The cells were washed three times and kept in fresh medium, then lysed with lysis/complementation buffer at 8 h post-infection. IN-ProLabel activity in the cell lysate was measured according to manufacturer's instructions from the assay kit (ProLabel™ Detection Kit II, Clontech Laboratories, Inc.)

3.2.13 Virus Production and Infection

To study the effect of Ku70-knockdown on HIV-1 replication, equal amounts (quantified by HIV-1 p24 Antigen) of pNL4.3-GFP virus were used to infect Ku70-knockdown or empty-vector-transduced C8166 T cells for 2 h; cells were then washed and cultured in a 37°C incubator. At different time points, viral replication levels were monitored by the measurement of p24 levels using an HIV-1 Gag-p24 ELISA. To test the infectivity of progeny virus produced from the Ku70- knockdown cells, empty-vector and Ku70-knockdown C8166 T cells were infected with the same amounts of pNL4.3-GFP. Progeny viruses were collected by ultracentrifugation after four days of infection and equal amounts of viruses (quantified by HIV-1 p24 Antigen) were used to infect empty-vector or Ku70- knockdown C8166 T cells. Viral infection was examined at three days

post-infection by monitoring HIV p24 levels in the supernatant.

3.2.14 Virus composition and incorporation of cellular protein into HIV-1 virion

To examine the viral protein compositions, the pNL4.3-GFP viruses from empty-vector-transduced and Ku70- knockdown C8166 T cells were pelleted through a 20% sucrose cushion at 35,000 rpm for 1.5 h at 4°C. Then, equal amounts of viruses (normalized by p24 values) were lysed with 4X Laemmli Buffer and directly loaded onto an SDS-PAGE gel and analyzed for IN and p24 expression using their corresponding antibodies. The reverse-transcriptase activity from the purified viruses was analyzed by Reverse Transcription Assay using a commercial RT assay kit (Roche Diagnostics) according to the manufacturer's instructions.

To detect the presence of Ku70 in the HIV-1 particles, 15 x 10^6 CD4+ C8166 T cells were mock infected or infected with pNL4.3-GFP for three days. Then, supernatants from both cell cultures were ultracentrifuged at 35,000 rpm for 1.5 h through a 20% sucrose cushion. The pellets were dissolved in the same volume of RIPA buffer and mixed with 20% v/v TCA, followed by precipitation on ice for 30 min and acetone washing. Protein precipitates were dissolved in 4X Laemmli Buffer and directly loaded onto a 10% SDS-PAGE gel. Virus- associated Ku70 and p24 were then examined by WB using the corresponding antibodies.

3.2.15 Subtilisin treatment of purified HIV-1 virions

The subtilisin assay was performed according to the protocol as described [293]. The Vpr-RT-IN or Vpr-RT expressor was cotransfected with VSV-G and NL4.3lucΔBglΔRI

to produce single cycle IN+ and IN- virus. The viruses were first ultracentrifuged through 20% sucrose at 35,000 rpm for 2 h and then mock treated or treated with 0.1 mg/ml of subtilisin (Sigma) for 20 h at 37°C incubator. Subtilisin was inactivated by phenylmethylsulfonyl fluoride. Virus was then repelleted as described above, lysed in RIPA butter and loaded onto SDS-PAGE followed by WB. Blots were sequentially probed with anti-Ku70, anti-IN and p24 antibodies.

3.2.16 Quantitative real-time PCR

To measure reverse transcription step of VSV-G pseudotyped single-cycle replicating viruses containing IN mutants EH170,1AA, EK170,3AA, and HL171,2AA, quantitative PCR (qPCR) analysis was conducted. Equal amounts of viruses (adjusted by virion-associated p24 levels) were treated with 340 IU/mL DNase (Roche Molecular Biochemicals) for 1 h at 37 °C to remove residual plasmid DNA and then used to infect C8166 CD4+ T cells. For negative control (NC), prior to DNase treatment, wt virus was inactivated by incubating at 70 °C for 0.5 h. The DNA was isolated from 1 x 10^6 C8166T cells at 12 h post-infection using QIAamp® DNA blood kit (Qiagen sciences, Maryland, USA) following the manufacturer's instruction. The reverse transcription activity of HIV-1 in the infected cells was analyzed by quantifying the total HIV cDNA by using the qPCR technique. The qPCR was performed on Mx3000P detection system (Stratagene, CA) using LightCycler FastStart DNA Master SYBR Green I master mix (Roche diagnostics, Germany) along with forward (5'-tac tga cgc tct cgc acc-3') and reverse (5'-tct cga cgc agg act cg-3') primers targeted to the 5' end of the LTR and Gag region of the HIV-1 Bru genome [294]. The optimized thermal conditions used in the qPCR were as follows: initial hot start (95 °C for 15 min) followed by 35 to 40 cycles of denaturation

(94 °C for 30 s), primer annealing (60 °C for 30 s) and extension (72 °C for 1 min). The total HIV-1 cDNA levels were expressed as copy numbers per cell, with DNA template normalized by the β-globin gene.

To examine the effects of Ku70 knockdown in the formation of late reverse transcription products (late RT), 2-LTR circles and integrated DNA, 1.5×10^6 stable C8166 T cell lines with Ku70- knockdown or empty-vector-transduced were infected with the pNL4.3-GFP virus as described above. Heat-inactivated virus (70°C for 30 min) was used as a negative control for infection. After 4 h infection, cells were washed and cultured in fresh RPMI medium. At 24 h post-infection, cells were harvested and washed with PBS twice. DNA was isolated using a QIAamp blood DNA minikit (Qiagen). The total levels of HIV-1 DNA, 2-LTR circles, and integrated DNA were quantified following the same procedure in an Mx3000P real-time PCR system (Stratagene, CA) as described [283].

3.2.17 Statistical analysis

To analyzing the data obtained from real-time PCR in chapter 6, student's *t* test was used to calculate the statistical significance, and a p value ≤ 0.05 was considered significant.

Chapter 4 (Published in Retrovirology 2008, 5:102)

Contribution of the C-terminal region within the catalytic core domain of HIV-1 integrase to yeast lethality, chromatin binding and viral replication

4.1 Rational

As discussed in Introduction section in chapter 1, IN has pleiotropic effects on various HIV-1 replication steps including reverse transcription, nuclear import, chromosomal targeting and integration. Thus, it is difficult to specifically study the effects of individual mutations on one particular function of IN during viral replication. Hence, other functional assays have been developed in order to elucidate the mechanisms underlying the biological activities of IN in eukaryotic cells. Several previous studies have reported a yeast eukaryotic system in which the expression of IN alone in some *S. cerevisiae* strains, such as the W303-1A rad52 mutant strain and the AB2 diploid strain, resulted in a lethal phenotype [295, 296]. These studies revealed that the IN-induced lethal phenotype may be related to the catalytic activity of IN as an IN catalytic mutant (D116A) was unable to induce the lethal phenotype in yeast [295, 297]. However, a subsequent study has reported that another mutant targeting E152 in IN, one of the three critical residues (D64, D116 and E152) essential for the catalytic activity of IN, did not disrupt the ability of IN to induce the lethal phenotype in yeast cells [298]. Therefore, the mechanisms for the IN-induced lethal phenotype in yeast still remain to be defined and it is likely that other functions of IN might be implicated in inducing the lethal phenotype in *S. cerevisiae.* A recent study reported that the expression of IN in yeast could mediate the integration of DNA containing viral LTRs into the yeast genome [299]. Given that

chromatin binding by IN is a prerequisite step for viral DNA integration [147, 154, 168, 300], it was hypothesized that the yeast lethal phenotype induced by IN might be due to its chromatin binding activity.

This project was initiated by a previous graduate student Zaikun Xu in Yao's lab, who generated a series of IN mutations and identified several residues (V165, A179, KR186,7) located in the C-terminal region of HIV-1 IN CCD that are required for the IN-induced lethal phenotype. He also found that these IN mutants were impaired in their ability to associate with cellular chromatin in 293T cells. My contribution to this project was to explore the potential involvement of cellular cofactors of IN including LEDGF/p75 and Ini1 in this process and to investigate the chromatin-binding ability of these IN mutants in yeast cells.

Two IN's cofactors including LEDGF/p75 and Ini1 might account for the impairment of chromatin-binding and lethal phenotype of IN mutants V165, A179, KR186,7AA. LEDGF/p75 is well characterized as the IN-to-chromatin tethering factor [147]. Another cellular co-factor Ini1 was originally discovered in a yeast two-hybrid system screening for cellular proteins interacting with IN [153]. Ini1 is a subunit of the SWI/SNF chromatin-remodeling complex [302] and it has been shown to increase the efficiency of integration in an *in vitro* assay [153]. Further studies have also found that Ini1 is capable of being incorporated into the HIV-1 virion and can modulate reverse transcription and Tat-mediated transcription [44, 286, 303]. Since the counterpart of Ini1 in yeast SNF5 is required for lethal phenotype induced by IN in yeast [301], it was hypothesized that these lethal phenotype- and chromatin-binding defect of these IN mutants (V165, A179, KR186,7AA) might result from impaired IN-Ini1 interaction.

4.2 Hypothesis

The hypothesis of this section is that lethal phenotype induced by HIV-1 IN in yeast requires its chromatin association and cofactor(s) binding ability.

4.3 Objectives

1. Characterize the mechanism(s) underlying the IN-induced lethal phenotype in yeast by testing its chromatin-binding and cofactors-binding ability.
2. Determine the importance of the C-terminal region of HIV-1 IN CCD in the viral life cycle.

4.4 Results

4.4.1 Effects of IN mutations on the lethal activity in HP16 yeast cells

Caumont et al. first demonstrated that HIV-1 IN induced a lethal phenotype in some yeast strains, including the JSC 302, W839-5C and AB2 strains, but not in the W303-1 strain (Rad52+) [296]. The system of the lethal phenotype induced by HIV-1 IN in the *S. cerevisiae* strain HP16 (*MATa ura3-52; his3Δ1; leu2; trp1Δ63; prb1-1122; pep4-3 prc1-407)* was successfully established in the lab. A yeast expression plasmid encoding the HIV-1 IN cDNA under the control of the galactose-inducible GAL1 promoter (p424Gal1-IN) was constructed and transformed into HP16 yeast cells which were cultured in inducible media (Trp^-, 2% galactose (Gal)) to drive the expression of IN. The empty vector was used as control.

After we confirmed that the expression of HIV-1 IN in the HP16 yeast strain is able to

induce the lethal phenotype (data not shown), we moved on to identifying the critical amino acid(s) or motif(s) in IN important for the induction of the lethal phenotype. Various IN mutants, including F1A, K136A, K159P, V165A, A179P, KR186,7AA, KK215,9AA and RK263,4AA, were introduced into HP16 yeast cells. Most of these mutants have been previously shown to disrupt HIV-1 replication at different steps, including proviral DNA integration [79, 112, 282, 304-306]. As determined by an anti-IN IP and WB, transformation of plasmids encoding these IN mutants resulted in comparable IN expression (Figure 6A lanes 2-10). Endogenous yeast β-actin was used as internal control. The effect of each mutant IN on yeast growth was measured in both liquid media and agar plates by the "drop test". Interestingly, our steady state analyses revealed that in the inducible media, yeast cells transformed with different IN mutants showed varying growth (Figure 6B and C). In particular, cells transformed with the IN mutants V165A, A179P or KR186,7AA had growth rate similar to the yeast cells transformed with empty vector, indicating that these three mutations, all located in the C-terminal region of the IN CCD, are unable to induce the lethal phenotype in the HP16 yeast strain. As such, we designated them as lethal phenotype-defective IN mutants. Therefore, these data indicate that the IN-induced lethal phenotype in HP16 yeast cells is not related to the catalytic activity of IN. This suggests that other functions of IN, which were affected by each of these three mutations, may play an important role for IN lethality in HP16 yeast strain.

4.4.2 Lethal phenotype-defective IN mutants are unable to efficiently associate with host chromatin.

While the IN mutants identified in this study do not induce the lethal phenotype in HP16

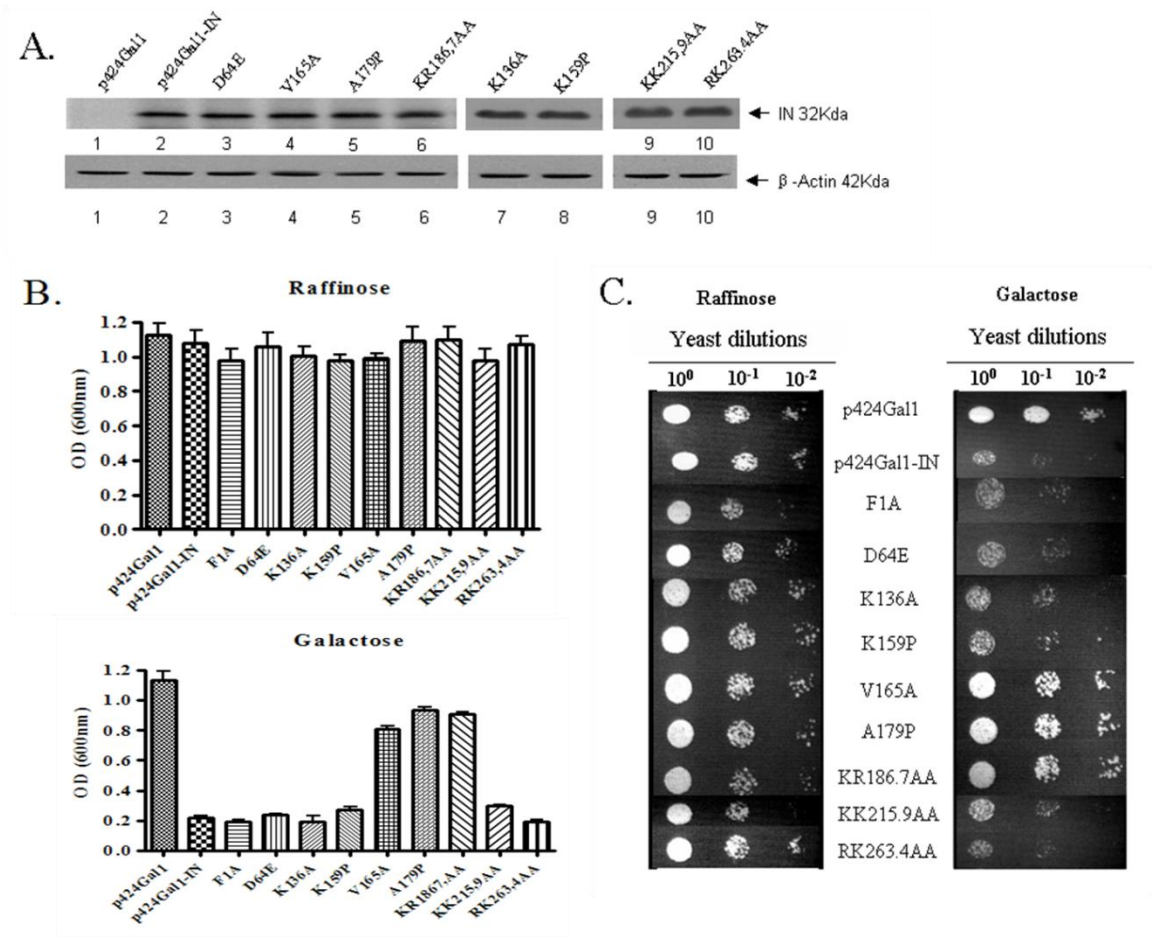

Figure 6 Effects of different IN mutants on yeast growth. A. Expression of different IN mutants in HP16 yeast cells. And the endogenous yeast β-actin was detected by anti-actin WB (lower panel). B. Each transformed yeast population was first grown in non-inducible selective media overnight. Then equal amounts of transformed yeast cells were grown either in non-inducible media (upper panel) or in the inducible media (lower panel) at 30°C for 24 h. Yeast growth was monitored by measuring each yeast cell culture density. Means and standard deviations from three independent experiments are shown. C. The growth of yeast in the absence of IN expression (left panel) or in the presence of IN expression (right panel) was also tested by the "drop test" on agar plate, as described.

cells, the mechanisms underlying this loss of lethality remained unknown. A recent study has demonstrated that the expression of IN in yeast could catalyze the integration of DNA containing the LTRs of HIV into the yeast genome with the same specificity in yeast and human cells [299]. This suggests that IN may utilize similar cellular machinery, including proteins important for chromatin tethering, for the integration process in yeast and mammalian cells. Therefore, we tested whether the lethal phenotype-defective mutants might affect the association of IN with chromatin by a chromatin binding assay, as previously described [289]. Yeast cells expressing either wild type or mutant IN were spheroplasted and lysed in a buffer containing 1% Triton X-100 for 5 min. The whole cell extracts were then incubated with 50mM or 200mM NaC1 for 20 min and the chromatin-bound and non-chromatin-bound fractions were separated, as described in the Materials and Methods. An anti-IN WB of chromatin-bound and non-chromatin-bound fractions showed that the wild type IN was exclusively detected in the chromatin-bound fraction, whereas the chromatin binding of the three lethal phenotype-defective IN mutants was impaired to differing degrees. Approximately 10% of the KR186,7AA IN bound chromatin and the association of the V165A and A179P IN with chromatin was also reduced to 70% of wild type (Figure 7A, left panel). As a control, the localization of the yeast homocitrate synthase isoenzymes Lys20/21p [307] was also evaluated and, as expected, Lys20/21p were exclusively associated with non-chromatin fraction (Figure 7A, right panel).

In order to determine whether these IN mutations also affected their chromatin binding in mammalian cells, we tested the binding of the IN mutants to chromatin in 293T cells using a chromatin isolation protocol as described previously [168]. As a control, the YFP

protein and the nuclear pore complex-associated protein Nup62 were also transfected into the 293T cells and analyzed for their chromatin association. Results showed that up to 20-25% of the wild type IN and mutants D64E and K136A were detected in the chromatin-bound P1 fraction (Figure 7B, left panel, lanes 2 to 4). Similar to what was observed in yeast cells, the V165A, A179P and KR186,7AA mutants were exclusively present in the non chromatin-bound S1 fraction (Figure 7B, left panel, lanes 5 to 7). As expected, the expressed YFP was only detected in the S1 fraction, whereas the Nup62 was detected in P1 fraction (Figure 7B, right panel). In order to confirm that the wild type IN and the D64E and K136A mutants were indeed associated with the chromatin, the chromatin-bound P1 fractions were further treated with DNase and salt to release the chromatin-bound proteins into the soluble fraction (S2). After being treated with DNase and salt, all of the chromatin-bound IN protein was released into S2 fraction (Figure 7B, left panel, lanes 2 to 4). However, the Nup62 protein remained in the insoluble fraction (P2) (Figure 7B, right panel). These results indicate that, similar to that in yeast, the lethal phenotype-defective IN mutants were impaired for their association with the cellular chromatin in mammalian cells.

To exclude the possibility that the loss of the ability to bind chromatin by these lethal phenotype-defective IN mutants is caused by a defect in nuclear translocation, we analyzed intracellular localization of different IN mutants by immunofluorescence in COS-7 cells. To avoid passive diffusion of the relatively low molecular weight IN protein into the nucleus, we constructed YFP-IN fusion proteins by fusing each IN mutant to YFP. An IN deletion mutant, YFP-IN1-212, which was previously shown to be unable to be localized in the nucleus [282], was used as a negative control. In contrast to

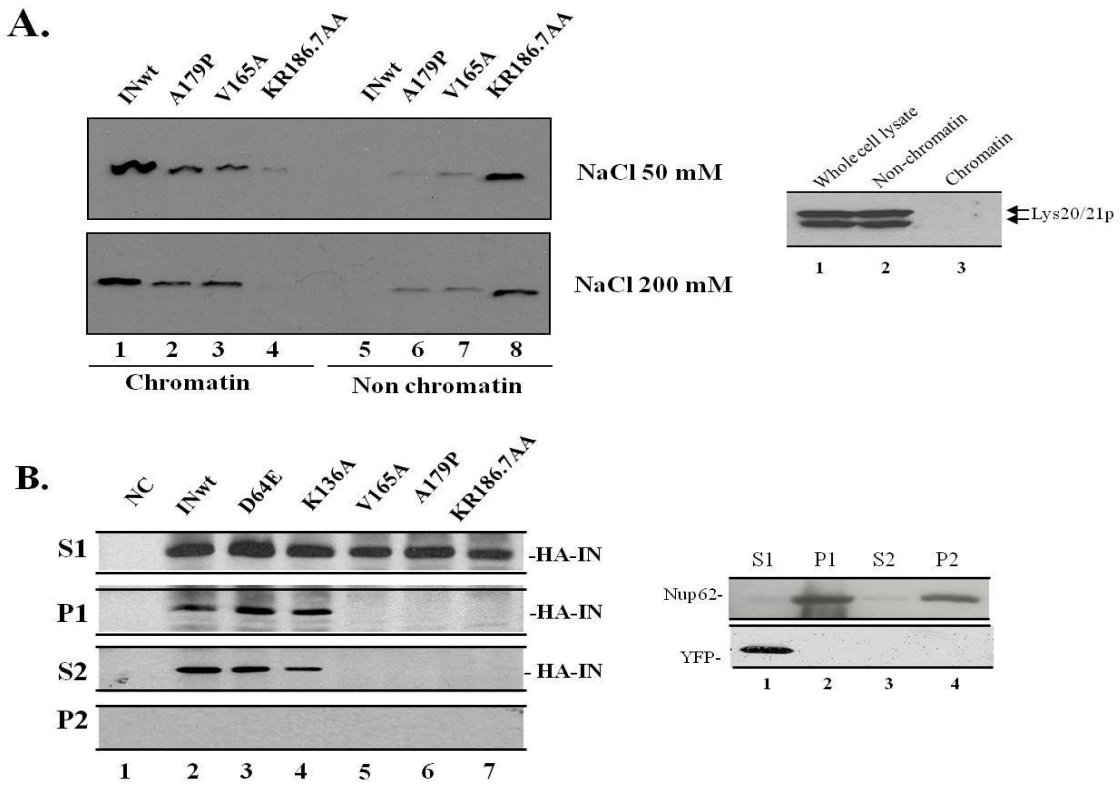

Figure 7 The lethal phenotype-defective IN mutants lack chromatin binding ability in yeast and mammalian cells. A. Chromatin binding ability of IN in yeast cells. After growing in IN-inducible media (Trp-, 2%gal+) for 3h, p424-Gal-IN wild type-, A179P-, V165A- and KR186,7AA-transformed yeast cells were lysed. Whole cell extracts were incubated with 50 or 200mM NaCl for 20 min before the separation of chromatin-bound and non-chromatin-bound fractions. Both fractions were subjected to WB using anti-IN antibody (left panel). Right panel: The wild type IN-expressing yeast cells were fractionated into chromatin- and non-chromatin-bound fractions and the Lys20/21 protein in each fraction was detected by WB using an antibody against Lys20.21. B. 293T cells were transfected with different SVCMV-HA-tagged-IN expressors (including the wild type IN and different mutants, as indicated. After 48h, transfected cells were lysed in cold CSK I buffer (0.5% Triton X-100), fractionated and the presence of HA-IN in the chromatin-

bound and non-chromatin-bound fractions was analyzed by IP and WB with anti-HA antibodies (left panel). In parallel, the presence of the nuclear pore complex-associated protein Nup62 in different fractions of 293T cells or 293T cells transfected with a SVCMV-YFP expressor was also analyzed using following the same procedure and by using anti-Nup62 and anti-GFP antibodies (right panel). S1: Supernatant (non-chromatin-bound fraction); P1: Pellet (chromatin-bound fraction); S2: DNase-released chromatin-associated proteins in P1; P2: insoluble, cytoskeletal, and nuclear matrix proteins in P1.

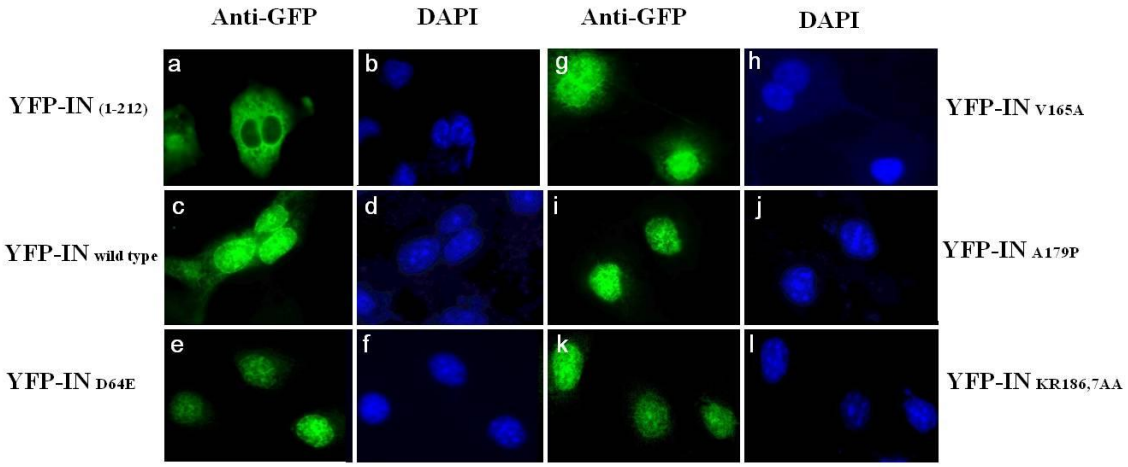

Figure 8 Intracellular localization of different IN mutants**.** COS-7 cells were transfected with different SVCMVin-YFP-IN fusion protein expressors as indicated. Cells were incubated with primary rabbit anti-GFP antibody followed by secondary FITC-conjugated anti-rabbit antibodies and the nuclei were stained with DAPI. Cells were visualized on a Carl Zeiss microscope (Axiovert 200) with a 63x oil immersion objective.

the cytoplasmic distribution of YFP-IN1-212 (Figure 8, a and b), the wild type and all of the IN mutant fusion proteins were predominantly localized in the nucleus (Figure 8, c to l). These results clearly indicated that yeast lethal phenotype-defective mutants retained the ability to translocate into the nucleus.

4.4.3 Differential binding of IN mutants to Ini1 or LEDGF/p75

Ini1/hSNF5 is a component of the chromatin remodeling SWI/SNF complex and was first identified as an interacting partner for HIV-1 IN [153, 302]. A previous study by Parissi et al. has shown that the inactivation of the SNF5 gene in yeast abolished the IN-induced lethal phenotype [301]. In order to determine if the failure of the IN mutants to induce the lethal phenotype may be due to their inability to bind to Ini1, the counterpart of SNF5 in mammals, we analyzed the interaction between the IN mutants and Ini1 using a cell-based Co-IP assay. After co-transfection with each IN-YFP expressor and pCGN-HA-Ini1 expressor into 293T cells, the binding of IN-YFP to HA-Ini1 was analyzed by an anti-GFP IP followed by an anti-HA WB. Unlike YFP alone (Figure 9A lane 2), IP of all IN-YFP fusion proteins, including the wild type IN and the three lethal phenotype-defective mutants, were able to co-precipitate similar amounts of Ini1 (Figure 9A. lanes 3-6). This suggests that mutations introduced at amino acids V165, A179 and KR186,7 did not affect their ability to interact with Ini1. Also, the total amount of HA-Ini1 in each sample was evaluated by an IP with an anti-HA antibody followed by an anti-HA WB analysis. Similar levels of HA-Ini1 were expressed in each sample (Figure 9A, lower panel).

Another known cellular protein that interacts with IN, LEDGF/p75, has been shown to

be a tethering factor that links IN to the chromatin during the early stage of the viral replication [154, 168, 300]. To test whether IN mutants that have impaired chromatin-binding ability also have altered binding to LEDGF/p75, we tested the interaction between IN and LEDGF/p75 by using the same Co-IP assay in 293T cells. The SVCMV-IN-YFP expressor and a SVCMV-T7-LEDGF expressor were co-transfected into 293T cells. Cells were lysed at 48 h post-transfection and the interaction of IN-YFP and T7-LEDGF/p75 was analyzed by an anti-GFP IP followed by an anti-T7 WB. Our results confirmed the specific interaction between the IN and LEDGF/p75, since only the wild type IN-YFP, not YFP alone, was able to co-precipitate T7-LEDGF/p75 (Figure 9B, compare lane 3 and lane 2). Interestingly, all three IN mutants (V165A, A179P, KR186,7AA**)** which lost their chromatin-binding ability failed to interact with LEDGF/p75 (Figure 9B, Lanes 4-6). To rule out the possibility that the differences in binding for each mutant may be due to different expression levels of T7-LEDGF/p75 in the transfected cells, each cell lysate was further evaluated with an IP using an anti-T7 antibody followed by anti-T7 WB. Similar levels of T7-LEDGF/p75 were expressed in each population of transfected cells (Figure 9B, lower panel). Together, these data indicate that chromatin binding- impaired mutants are unable to bind to LEDGF/p75.

4.4.4 HIV-1 viruses encoding the lethal phenotype-defective IN mutations are replication defective

In order to characterize the effect of these lethal phenotype-defective IN mutants on HIV replication, we introduced each of these IN mutations into a previously described single-cycle HIV-1 replication system and evaluated viral replication [282]. After production of each virus stock, the virion-incorporated RT, IN and Gag were analyzed by WB using an

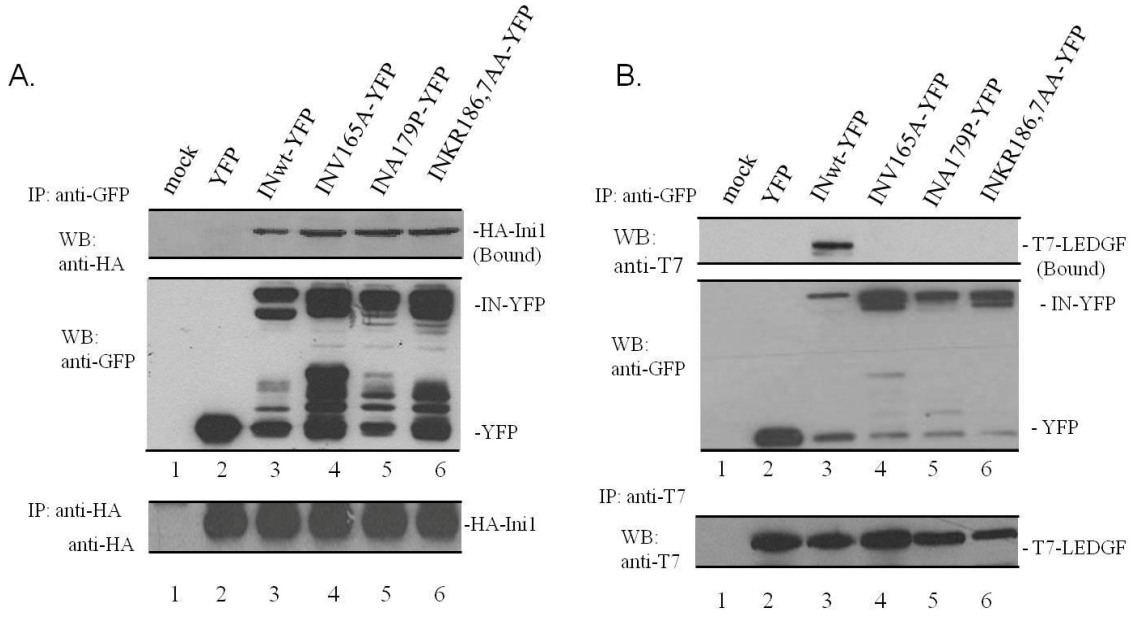

Figure 9 Characterization of IN mutants binding to Ini1 and LEDGF/p75. A. Interaction of wild type or mutant IN with Ini1. SVCMV-IN-YFP plasmids expressing wild type or mutant IN were co-transfected with pCGN-HA-Ini1 into 293T cells. The CMV-YFP expressor was used as a control. At 48 h post-transfection, cells were lysed and immunoprecipitated with rabbit anti-GFP antibody and subjected to WB with anti-HA antibody to measure the amount of co-precipitated Ini1 (upper panel). The same membrane was stripped and blotted with anti-GFP antibody to detect IN-YFP and YFP expression (middle panel). The unbound Ini1 was also checked by sequential IP with mouse anti-HA antibody followed by WB with the same antibody (lower panel). B. Lethal phenotype-defective IN mutants do not bind to LEDGF/p75. The SVCMV-IN-YFP plasmid expressing wild type or mutant IN or the SVCMV-YFP plasmid was co-transfected with SVCMVin-T7-LEDGF expressor in 293T cells. Cells were lysed 48 h post- transfection, whole cell protein extracts were immunoprecipitated with rabbit anti-

GFP followed by a WB using mouse anti-T7 HRP-conjugated antibody to detect the co-precipitated T7-LEDGF (upper panel). Also, the expressions of IN-YFPs were detected using an anti-GFP HRP-conjugated antibody (middle panel). The unbound T7-LEDGF was also checked by sequential IP with anti-T7 antibody followed by a WB with the same antibody (lower panel).

anti-HIV serum. Each IN mutant virus contained similar levels of IN, RT, and Gag proteins, compared to the wild type virus, indicating that incorporation of RT and IN, as well as HIV-1 Gag processing, was not affected by introducing various IN mutations (data not shown).

To test the infectivity of the IN mutant viruses, we infected the C8166 $CD4^+$ T cell line with equal amounts of VSV-G-pseudotyped IN mutant viruses. Since all IN mutant viruses contained a luciferase gene in place of the *nef* gene, viral infectivity was monitored by using a sensitive luciferase assay, as described previously [282]. While the wild type IN virus infection resulted in a high level of luciferase activity and peaked (1.5 $x10^5$ RLU) at 64 h post-infection in dividing C8166 cells (Figure 10). The infection of the class I mutant D64E virus only resulted in a basal level of luciferase activity that was approximately 10^4-fold lower than the wild type (Figure 10). Interestingly, when C8166 cells were infected with VSV-G-pseudotyped viruses containing the chromatin binding-defective IN mutants, the levels of luciferase activity were similar to the D64E mutant virus throughout the 6-day period (Figure 10). These results indicate that, like the class I D64E mutant virus, the chromatin binding-deficient V165A, A179P and KR186,7AA mutant viruses were replication defective in C8166 cells.

To determine whether the replication defect in IN mutant viruses could be due to a defect at the reverse transcription level, we analyzed viral DNA synthesis following infection of C8166 cells with each IN mutant. The levels of late reverse transcription products were analyzed by semi-quantitative PCR at 12 h post-infection with HIV-1-specific 5'-LTR-U3/3'-Gag primers [282]. Total viral DNA synthesis during infections with the V165A-, A179P- and KR186,7AA-containing viruses was similar to that following infection with

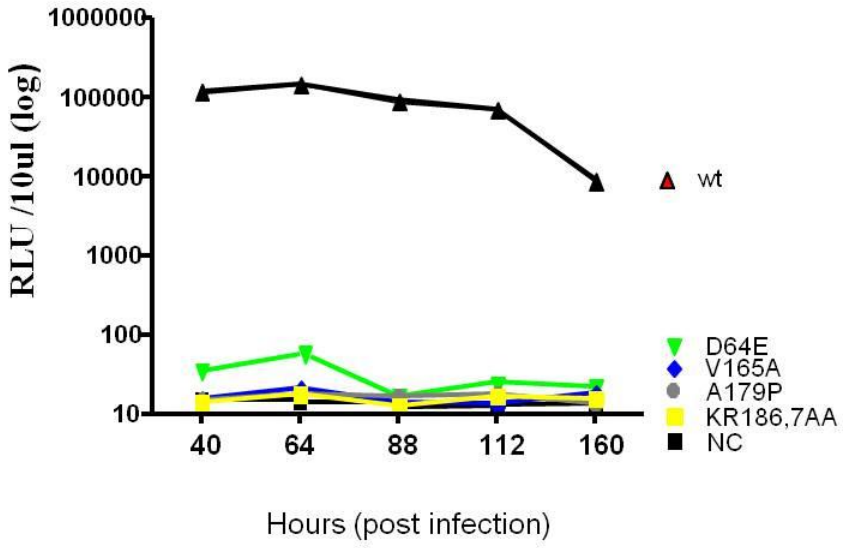

Figure 10 Effects of lethal phenotype-defective IN mutants on VSV-G-pseudotyped HIV-1 replication. To assess viral infection, equal amounts of VSV-G-pseudotyped virions containing various IN mutants were used to infect the CD4+ C8166 T cell line. At different time points, equal cell numbers were collected and the infection generated by each virus stock was evaluated by luciferase assays. The results are representatives of three independent experiments. NC: negative control.

the wild type and D64E mutant viruses (data not shown). These results indicate that all three lethal phenotype-defective IN mutants did not significantly affect reverse transcription during single-cycle replication of HIV-1.

4.5 Discussion

4.5.1 Lethal phenotype in yeast and chromatin binding ability of IN

HIV-1 IN plays a critical role in several steps of the early viral replication, including reverse transcription, nuclear import of viral DNA and integration [16, 90, 92, 149, 282, 308]. In order to further investigate different functions of HIV-1 IN and the molecular mechanisms involved, numbers of *in vitro* and *in vivo* assays, including a yeast IN expression system, have been developed to specifically assess the different activities of HIV-1 IN. Caumont et al. initially reported that the expression of HIV-1 IN in some yeast strains resulted in a lethal phenotype [296]. However, the mechanism for the IN-induced lethal phenotype in yeast still remains to be fully understood. In this study, we introduced different IN mutants into the HP16 strain of *S. cerevisiae* and tested their effects on yeast viability. Three IN mutants (V165A, A179P and KR186,7AA) were identified as lethal phenotype-defective mutants in HP16 cells (Figure 6).

A recent study has demonstrated that the expression of IN in yeast could mediate the integration of DNA containing viral LTRs into the yeast genome [299]. Thus, it appears that the molecular mechanisms underlying the activity of IN in both yeast and mammalian cells are similar. Given that chromatin targeting by IN is a prerequisite step for viral DNA integration [147, 154, 168, 300], we investigated whether the yeast lethal phenotype-defective mutants would also have a defect in chromatin binding.

Interestingly, the chromatin binding experiments showed that three lethal phenotype-defective mutants were impaired for binding to chromatin in both yeast and mammalian cells. The reduced chromatin binding ability of the IN mutants did not appear to result from the diminished nuclear entry of IN as we found that the nuclear translocation of these IN mutants was intact. Overall, these data clearly indicate that these mutations at the C-terminal region of the IN CCD severely affected its ability to interact with host cell chromatin.

4.5.2 Chromatin binding and cofactors binding of HIV-1 IN

We further investigated the possible mechanism underlying this loss of chromosomal binding by testing the interactions of these IN mutants with two known IN-interacting cellular proteins. The first, Ini1, is a homolog to the yeast SNF5 protein and was found to bind to HIV-1 IN in a yeast two-hybrid system and was shown to increase the efficiency of integration in an *in vitro* assay [153]. Also, inactivation of the SNF5 gene has been shown to abolish the IN-induced lethal phenotype in yeast [153]. However, our results indicate that the three yeast lethal phenotype-defective mutants were able to bind Ini1 at a level comparable to the wild type IN (Figure 9A), indicating that the loss of lethality in these IN mutants is not related to IN-Ini1 binding ability. However, at this point, we still cannot rule out the possibility that SNF5 may act on other unidentified steps to affect the IN-induced lethal phenotype.

The second cellular protein examined, LEDGF/p75, has been previously shown to directly interact with HIV-1 IN [147]. LEDGF/p75 functions in the tethering of IN to the host chromosome, a step essential for HIV-1 integration and viral replication [154, 168,

300]. Interestingly, we found that all three lethal phenotype-defective IN mutants did not interact with LEDGF/p75 (Figure 9B) in 293T cells. Previous study have identified two regions within IN that are involved in the interaction with LEDGF/p75: the region around W131 and W132 and the region from I161 to E170 [173]. In line with this observation, the structural analysis of IN suggests that α1 and α3 helices in the B chain of IN, in addition to the α4/α5 connector (residues 166-171) and α5 helix in the A chain form the interface of the LEDGF/p75 binding site [169, 171]. These studies also indicated that the V165A and L172A mutants were unable to binding LEDGF/p75 *in vitro*. In this study, we confirmed that the V165A mutant was indeed unable to interact with LEDGF/p75. Moreover, two other IN mutants (A179P and KR186,7AA) were identified in this study that were incapable of binding to both LEDGF/p75 and to chromatin. As both mutations are located in the region encompassing amino acids from 171 to 186, which has been predicted to form the putative helix-α5 structure and/or the α5/α6 connector (residues 166-171) [261], it suggests that these regions and/or the helical structure play an important role in the interaction between IN and LEDGF/p75 and its chromatin tethering function. However, an important question that needs to be addressed is how this region contributes to the interaction between IN and LEDGF/p75. Interestingly, a recent study by Berthoux et al. demonstrated that the lysine at position 186 is critical for IN multimerization. However, their study also indicated that the K186Q mutant interacted with LEDGF/p75 as efficiently as the wild type IN in a two-hybrid assay [79]. In contrast, another recent report by Mckee et al. confirmed that mutations introduced in the KRK(186-188) region affected IN multimerization. Furthermore, their study also indicated that this KRK(186-188) region is important for

IN binding to LEDGF/p75 [309]. Since our KR186,7AA mutant lost the ability to bind to LEDGF/p75 and host chromatin, it still remained unknown if our other lethal phenotype-deficient IN mutants could also affect multimerization of IN. Interestingly, it has been reported that there is no yeast homolog of LEDGF/p75 [299]. Therefore, it is tempting to speculate that there may be other host protein(s) in yeast that contribute to chromatin-targeting of IN. In this regard, there is another cellular protein, Hsp60, has also been shown to be important in IN-induced lethality in yeast [310]. However, the role that Hsp60 plays in the chromatin targeting and/or other activities of IN in yeast still awaits further characterization.

4.5.3 The effect of yeast lethal phenotype-defective mutants on HIV-1 replication

The effect of these yeast lethal phenotype-defective mutants on HIV-1 replication was also evaluated using a previously described single cycle infection system with VSV-G-pseudotyped HIV. Similar to the IN class I mutant D64E, all yeast lethal phenotype-defective IN mutant viruses were replication deficient (Figure 10). Since these IN mutants lost chromatin-binding and LEDGF/p75-binding abilities, the infection data highlighted the importance of chromatin association and LEDGF/p75 binding in IN functions and in viral replication. Overall, this study has established a functional correlation between the IN-induced lethality in yeast and the inability of IN to associate with host chromatin and bind to LEDGF/p75. This suggests that this yeast-based IN expression system may be a valuable system for studying the molecular mechanisms underlying the chromatin binding activity of HIV-1 IN, as well as for the high-throughput screening anti-IN molecules that specifically prevent IN from associating with chromatin. Also, our study suggests that a region encompassing amino acids 171 to

186, which was previously predicted to form the alpha-helix-5 structure [261], may play an important role in the binding of IN to both chromatin and LEDGF/p75. A more detailed analysis will be required to fully elucidate how this region and/or the secondary structure of IN contribute to this particular function of IN during viral replication.

Chapter 5 (Published in Virology Journal 2010, 7:68)

Characterization of the HIV-1 Integrase Chromatin- and LEDGF/p75-binding Abilities by Mutagenic Analysis Within the Catalytic Core Domain of Integrase

5.1 Rational

A number of previous studies have employed *in vitro* biochemical approaches to study the interaction between IN and DNA substrates by using oligonucleotides that mimic the HIV-1 LTR, and they have identified several residues in the IN that are responsible for binding viral DNA [311, 312]. Indeed, all three domains of IN, including the NTD, CCD and CTD, have been shown to interact with DNA by *in vitro* studies [313-315]. However, how IN interacts with host chromatin under physiological conditions is considerably less well understood.

Currently, the most widely accepted IN-to-chromatin model was proposed by Engelman. In this model, chromatin-associated LEDGF/p75 engages IN-bound viral DNA to a nearby genomic locus to carry out strand transfer reaction ; when LEDGF/p75 is reduced or depleted, integration of viral DNA will rely on less efficient cofactor-independent pathway or utilize alternative tethering factor such as HRP2 for integration [176]. This model has been supported by ample evidence [120, 140, 177, 178]. For example, HIV-1 integration in LEDGF/p75 knockout cells still remained 11% of the wild type level, suggesting that although LEDGF/p75 is important for IN-to-chromatin tethering and integration , it is not strictly essential [140]. HRP-2 knockdown in the human somatic LEDGF/p75 knockout cell line further reduced HIV-1 integration and replication [177]. Despite of these observations, the clear architecture of the functional IN/LEDGF/DNA

complex as well as the way in which both IN and LEDGF/p75 interact and work on both the viral DNA and host chromatin in the process of integration remain elusive. In chapter 4, we have identified three IN mutations (V165A, A179P, KR186,7AA) that impaired binding to host chromatin also failed to interact with LEDGF/p75 by using a cell-based chromatin binding assay and co-immunoprecipitation (Co-IP) [292]. Further mutagenic analysis for IN/chromatin and IN/LEDGF interactions may not only help to elucidate the molecular mechanism of the IN/chromatin tethering and binding but also facilitate the identification of novel cellular factor(s) involved in this important viral replication step.

5.2 Hypothesis

The hypothesis of this section is that LEDGF/p75 is not strictly required for IN binding to chromatin.

5.3 Objectives

1 Characterize chromatin- and LEDGF/p75- binding phenotypes of a series of IN mutants

2 Investigate LEDGF/p75-independent chromatin binding of IN

3 Determine the relative importance of chromatin/LEDGF binding abilities of IN in HIV-1 infection

5.4 Results

5.4.1 Analysis of different HIV-1 IN mutants for their chromatin- and LEDGF/p75-binding.

Our previous study showed that three IN CCD mutants V165A, A179P, KR186,7AA, which cannot bind LEDGF/p75, lack the ability to bind to host chromatin (chapter 4) [292]. In the present study, we carried out a detailed mutagenic analysis to define binding site(s) for chromatin and LEDGF/p75 within the CCD of IN. Besides the previously reported IN mutants, V165A, A179P, KR186,7AA and a class I mutant D64/D116AA [292], several new YFP-IN mutants were generated by site-directed mutagenesis. The region E170-K173 was of interest because it overlaps with α-helices 4/5 connector residues 166–171 residing at the IN-LEDGF crystal interface [171]. Meanwhile, mutagenic studies have highlighted the importance of E170A, H171A, LK172,3AA for LEDGF/p75 interaction [169, 172, 173]. The mutants K136, K159 were also included as they were reported to be involved in IN/nucleotide binding [316-318]. To address the role of α-helix 6 of IN in chromatin- and LEDGF interaction, mutants I200A and I203A, I203P were also included in the study. Table 1 lists 17 IN amino acid residues analyzed in the study, their conservations in different HIV-1 isolates (the HIV sequences database was downloaded from the LANL website (http://www.lanl.gov) and aligned with MEGA4 program) and mutations introduced for each residue(s).

	Conservations *
Wild type	
D64/D116	99.4/99.7
K136	31.3
K159	99.5
V165	93.5
E170	99.6
H171	98.5
L172	99.4
K173	96.9
V176	99.4
A179	99.8
I182	98.0
F185	99.4
KR186,7	99.7/99.0
I200	98.3
I203	96.8

Mutations	Chromatin binding	Interaction with LEDGF/p75
Wild type	+++	+++
DD64,116EA	+++	++
K136A	+++	++
K159P	-	+/-
V165A	-	-
H171A	+++	-
L172A	+++	-
EH170,1AA	++	-
EK170,3AA	+++	+++
HL171,2AA	-	-
HK171,3AA	++	++
V176A	-	-
A179P	-	-
A179I	-	NA
I182A	+++	+++
F185A	+	++
KR186,7AA	-	-
I200A	+	-
I203A	+++	++
I203P	-	-

* Percent identity at that position among a collection of 1242 HIV-1 and SIV_{cpz} strains (http://www.hiv1.lanl.gov).

Table 1 Summary of IN mutant chromatin/LEDGF binding phenotypes. The percentage of selective amino acid conservation of IN is illustrated in the left panel, while chromatin-binding and LEDGF/p75-binding abilities of these IN mutants are shown in the right panel.

These IN mutants were further subjected to the chromatin binding assay [168, 292, 319] to study their host chromatin binding abilities. Briefly, each of YFP-INwt/mut was transfected into 293T cells, and, after 48 h, the presence of each YFP-INwt/mut in chromatin- and non-chromatin–bound fractions were analyzed by WB with anti-GFP antibody, as described previously [292]. Our data showed that, in addition to the previously described IN mutants (V165A, A179P, KR186,7AA [292]), K159P, V176A, A179I, I203P were also severely impaired for host chromatin binding (Figure 11A, data not shown for A179I). By contrast, mutants K136A, H171A, L172A, I182A and I203A were still able to associate with chromatin. The chromatin binding affinity of F185A and I200A was reduced by approximately 60% of wild type IN (Figure 11A).

Because LEDGF/p75 has been shown to be involved in chromatin targeting of IN, we also tested the LEDGF/p75-binding ability of different IN mutants by a cell-based Co-IP assay. Equal amounts of T7-LEDGF and CMV-YFP-IN wt/mut plasmids were co-transfected into 293T cells. After 48 h of transfection, IN/LEDGF/p75 interaction was analyzed by Co-IP of YFP-IN with anti-GFP antibody followed by WB with anti-T7 antibody. Results revealed a strong interaction between T7-LEDGF and YFP-IN wild type and mutants D64E/D116A, K136A, I182A, F185A, I203A. Meanwhile, the mutants K159P, H171A, and I200A showed reduced affinity for LEDGF/p75 (Figure 12A, lanes 4, 6, and 13). Interestingly, several IN mutants including V165A, L172A, V176A, A179P, KR186,7AA, I203P lost their interaction with LEDGF (Figure 12A. lanes 5, 7, 8, 9, 12, and 15). As expected, no T7-LEDGF/p75 was pulled down by YFP control (Figure 12A, lane 1). To ensure that similar amounts of T7-LEDGF/p75 and YFP-IN were expressed in each sample, the presence of T7-LEDGF/p75 and YFP-IN in each sample was

detected by WB with corresponding antibodies (Figure 12A, middle and lower panel). The host chromatin and LEDGF/p75 interaction data of all the IN mutants analyzed in this study have been summarized in Table 1. Interestingly, we noted that IN mutants, H171A and L172A, displayed a drastically reduced interaction with LEDGF/p75 but still retained the interaction with chromatin.

5.4.2 Chromatin- and LEDGF/p75-binding analysis of IN double mutants within Loop $_{170}$EHLK$_{173}$

Interestingly, two IN mutants H171A and L172A that showed differential binding abilities to chromatin and to LEDGF/p75 are located in the CCD loop region $_{170}$EHLK$_{173}$ of IN, a connector that links helices α4 and α5. Thus, we then focused our studies on this region, which may be important for LEDGF/p75-binding, but not for IN chromatin-association. Indeed, this region overlaps with the interface for LEDGF-binding in the crystal study [171], and some IN mutants within this region, such as E170A, H171A, and LK172,3AA, have been shown to be impaired in the ability to bind LEDGF/p75 [169, 172, 173]. To further elucidate the functional roles of loop $_{170}$EHLK$_{173}$ on its chromatin and LEDGF-binding, we characterized the binding affinities of this region by testing the double mutants YFP-IN EH170,1AA, HL171,2AA, EK170,3AA and HK171,3AA (Figure 13A). The chromatin-association experiment showed that three of the double mutants EH170,1AA, EK170,3AA and HK171,3AA displayed strong binding affinity with cellular chromatin, whereas HL171,2AA completely lost its chromatin binding ability (Figure 13B). Meanwhile, the LEDGF/p75-binding ability of each mutant was also tested by Co-IP assay, and results showed that all the mutants except YFP-IN EK170,3AA lost their ability to interact with LEDGF/p75 (Figure 13C). The differential

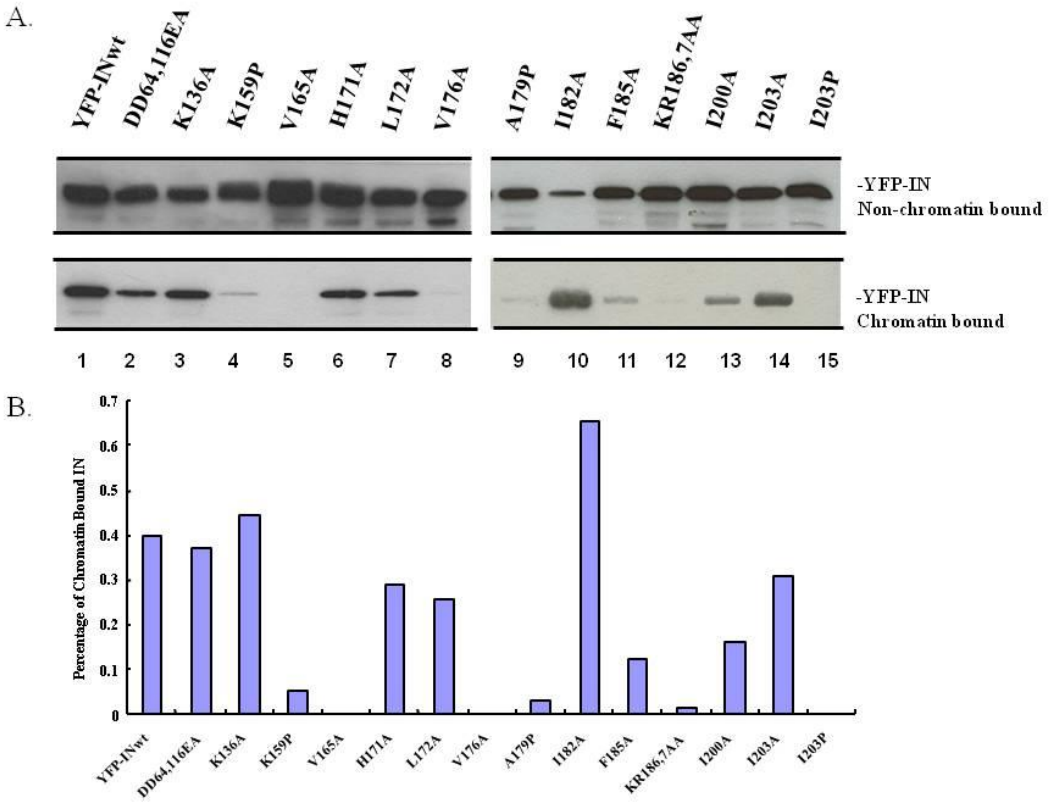

Figure 11 Identification of chromatin binding sites within IN CCD. A). 293T cells were transfected with different CMV-YFP-IN expressors (including the wild type IN and different mutants, as indicated). At 48 h post-transfection, cells were fractionated into chromatin-bound (lower panel) and non-chromatin-bound (upper panel) fractions as described in Material and Methods. YFP-IN in each fraction was analyzed by IP and WB with anti-GFP antibody. B). The intensity of both the chromatin-bound and non-chromatin-bound YFP-IN was densitometrically determined. The data are presented as the percentage of chromatin-bound YFP-IN to total input. Results are representative of two independent experiments.

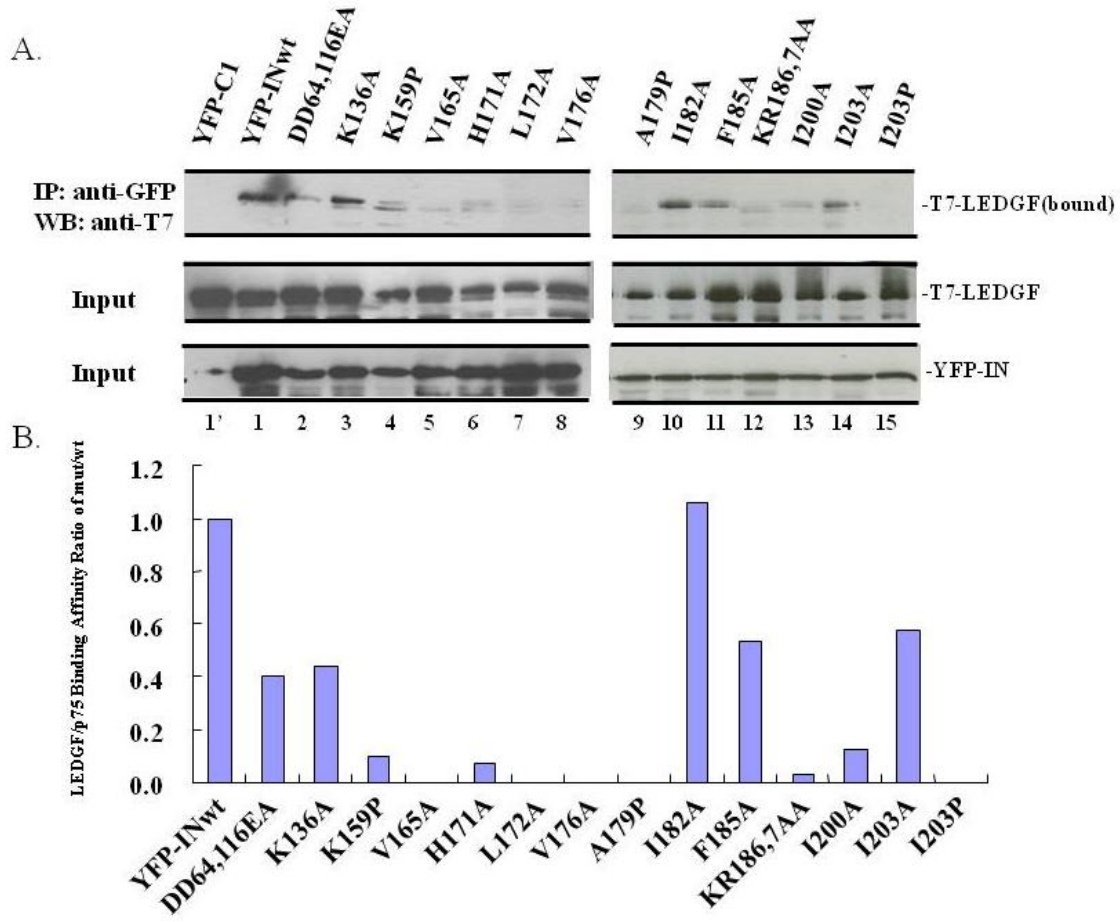

Figure 12 Identification of LEDGF/p75-binding sites within IN CCD. A). The CMV-YFP-INwt/mut or CMV-YFP plasmid was co-transfected with SVCMVin-T7-LEDGF expressor in 293T cells. After 48 h of transfection, 90% cells were lysed and subjected to Co-IP assay. The IN bound T7-LEDGF/p75 was precipitated by using rabbit anti-GFP antibody and detected by WB using mouse anti-T7 antibody (upper panel). 10% cells were lysed with 0.5% NP-40, directly loaded on 10% SDS-PAGE gel and probed with anti-T7 or anti-GFP antibody to detect T7-LEDGF or YFP-IN expression (middle or lower panel). B). The intensity of protein bands was densitometrically determined. Results were expressed as the ratio of bound T7-LEDGF/p75 expression (mutants/wild-type) which was normalized by total input. Binding affinity to LEDGF/p75 of YFP-IN

wild type was arbitrarily set as 100%. Results are representative of two independent experiments.

LEDGF-binding abilities of these four IN double mutants were re-confirmed by chemiluminescent Co-IP assay (Figure 13D). Altogether, uncoupled chromatin- and LEDGF-binding affinities were observed for IN mutants H171A, L172A and EK170,1AA, with strong binding affinity to chromatin but dramatically impaired contact with LEDGF/p75.

5.4.3 Nuclear localization of IN mutants in COS-7 cells

Since HIV-1 IN has been shown to be a karyophilic protein and is involved in nuclear import of PICs, we wondered whether introducing mutations in the $_{170}$EHLK$_{173}$ region of IN might interfere with IN nuclear translocation, which consequently affects their association with chromatin and/or LEDGF/p75 binding. To address this question, we transfected each IN mutant into COS-7 cells and analyzed their intracellular localization by immunofluorescence. Given the low expression of the YFP-IN fusion protein in COS-7 cells, the indirect immunofluorescence technique was used (as described in Materials and Methods). Results showed that, while the wild type IN was localized in the nucleus, the IN C-terminal deletion mutant YFP-IN1-212 was excluded from the nucleus, consistent with previous studies [292]. Also, all the IN $_{170}$EHLK$_{173}$ region mutants, including EH170,1AA, HL171,2AA, EK170,3AA and HK171,3AA, were able to accumulate predominantly in the nucleus (Figure 14). All of these results indicate that 1) the $_{170}$EHLK$_{173}$ region is dispensable for IN nuclear localization; and 2) the LEDGF/p75- and/or the chromatin-binding defects of those IN mutants were not due to their impaired nuclear translocation.

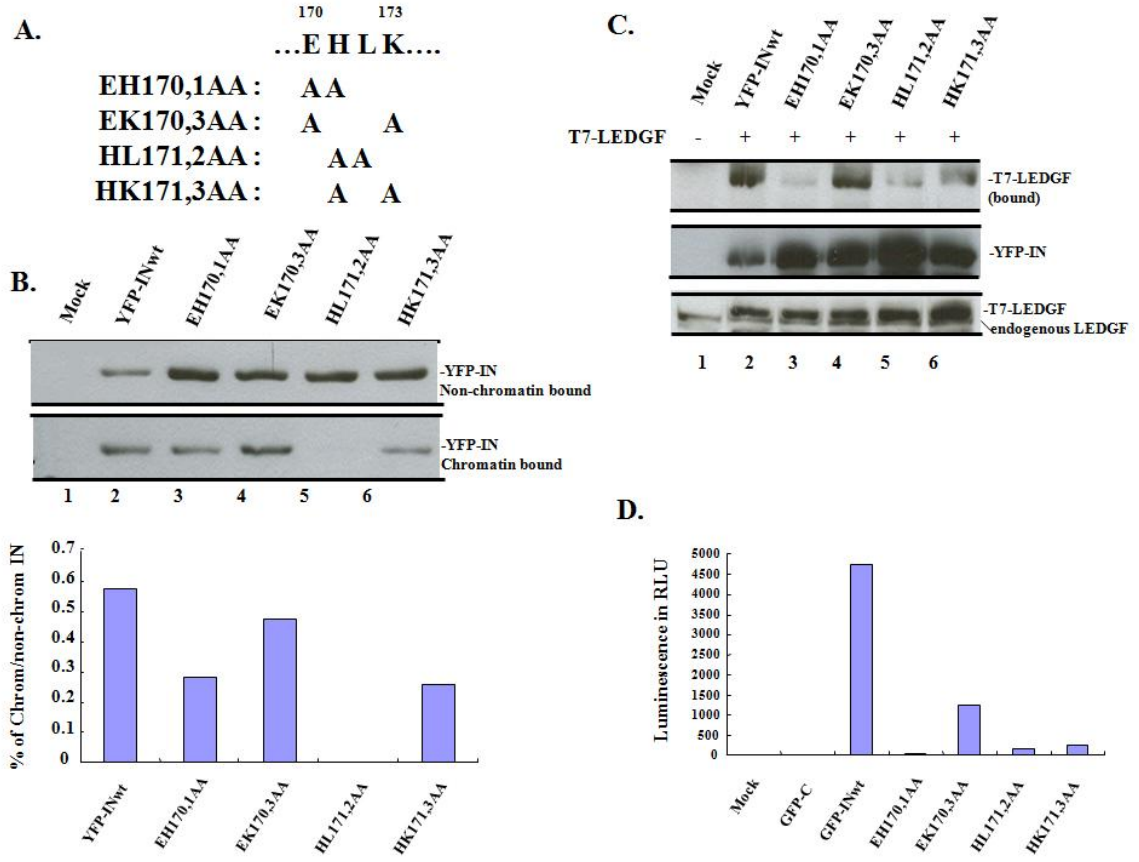

Figure 13 Differential effects of IN mutants within $_{170}$EHLK$_{173}$ region on chromatin- and LEDGF-binding. A). Diagram of amino acids sequence and introduced mutations in HIV-1 IN $_{170}$EHLK$_{173}$ domain. B). Chromatin binding profiles of IN double mutants within $_{170}$EHLK$_{173}$. 293T cells were mock-transfected or transfected with equal amount of CMV-YFP-IN wild type or double mutants EH170,1AA, EK 170,3AA, HL171,2AA and HK171,3AA. At 48 h post-transfection, cells were fractionated into chromatin-bound and non-chromatin-bound fractions as described in Material and Methods. YFP-IN in each fraction was analyzed by IP and WB with anti-GFP antibody. Chromatin binding affinity was quantified by laser densitometry and results are shown as the percentage of chromatin-bound to total input of YFP-IN (lower panel). C) LEDGF-binding affinity within IN $_{170}$EHLK$_{173}$ by Co-IP assay. 293T cells were co-transfected

with the SVCMVin-T7-LEDGF/p75 expressor and CMV-YFP-INwt/mut plasmid as indicated. After 48 h of transfection, 90% of cells were lysed and subjected to Co-IP assay as described before. The upper panel showed the bound T7-LEDGF/p75 in each sample. 10% of cell lysates were used to detect the expression of YFP-INwt/mut and T7-LEDGF/p75 by WB using anti-GFP and anti-LEDGF antibodies respectively (middle panel and lower panel). D). LEDGF-binding affinity within IN $_{170}$EHLK$_{173}$ detected by chemiluminescent Co-IP assay. AcGFP1-INwt/mut or AcGFP1-C and ProLabel-LEDGF fusion proteins were coexpressed in 293T cells. After 48 h of transfection, cells were lysed and immunoprecipitated with anti-GFP antibody and the chemiluminescent signals from ProLabel-LEDGF present in the complexes were measured by using ProLabel Detection Kit II and valued as relative luminescence units (RLU). Results are representative of two independent experiments.

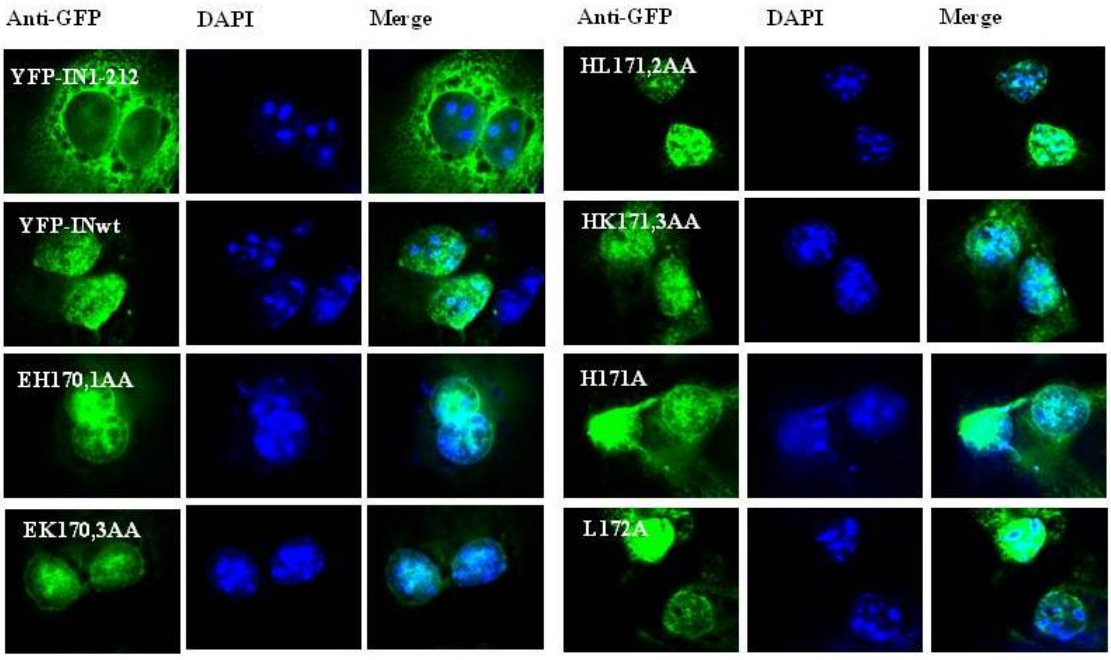

Figure 14 Subcellular localization of IN $_{170}$EHLK$_{173}$ mutants in COS-7 cells**.** COS-7 cells were transfected with different CMV-YFP-IN fusion protein expressors as indicated for 48 h. After fixation and permeabilization, cells were incubated with primary rabbit anti-GFP antibody followed by secondary FITC-conjugated anti-rabbit antibodies, and the nuclei were stained with DAPI. Cells were visualized by a Carl Zeiss microscopy (Axiovert 200) with a 63x oil immersion objective.

5.4.4 Knockdown of LEDGF/p75 had no effect on IN's chromatin binding

Uncoupled chromatin- and LEDGF-binding affinities observed in IN mutants within the $_{170}$EHLK$_{173}$ region suggest that LEDGF/p75 may not be essential for IN binding to chromatin. To gain more insight into the association between IN/chromatin binding and IN/LEDGF interaction, we tested the effect of LEDGF/p75 knockdown (LEDGF/p75-KD) on IN chromatin binding affinity. To obtain high efficiency gene knockdown, both synthetic siRNAs and shRNAs were combined in the study to knockdown LEDGF/p75 expression in 293T cells, as described in Materials and Methods. The results showed that such combined transient and stable LEDGF/p75-knockdown resulted in over 90% reduction of LEDGF/p75 expression (Figure 15B, lower panel). Then, the nuclear localization of HIV-1 IN in LEDGF/p75- knockdown cells was analyzed by indirect fluorescence using anti-LEDGF antibody. As shown in Figure 15A (lower panel), control cells transfected with scramble siRNA displayed abundant LEDGF/p75 protein expression. However, only a trace amount of LEDGF/p75 was detected in 293T cells transiently transfected with siRNA. Then, the cells were stained with anti-GFP antibody to visualize the localization of IN. Results showed that the wild type YFP-IN in transient LEDGF/p75- knockdown cells still accumulated in nuclei, suggesting that the LEDGF/p75- knockdown did not exert any significant effect on IN nuclear localization (Figure 15A, upper panel).

Next, we checked whether LEDGF/p75 depletion has an effect on IN chromatin binding. To do so, the LEDGF/p75-knockdown 293T cells were transfected with YFP-INwt, and after 24 h of transfection, cells were treated with MG-132, a proteasome inhibitor, to

prevent IN degradation. Cells were processed for IN chromatin binding analysis at 48 h post-transfection, as described above. Of note, no significant difference in the IN chromatin association was observed between the LEDGF/p75-knockdown cell line and the mock-transfected cell control (Figure 15B, upper panel). In parallel, the 293T cells transfected with the YFP-IN V165A mutant, which has been shown to be defective of chromatin binding, was used as a negative control [292]. Thus, our results demonstrated that the LEDGF/p75 knockdown could not abrogate IN chromatin binding.

5.4.5 Effect of IN $_{170}$EHLK$_{173}$ mutants on HIV-1 infection.

From the above results, we observed that LEDGF/p75 may not be mandatory for IN targeting to host chromatin. However, we still do not know whether LEDGF/p75-independent chromatin binding of IN could ensure HIV infection. To address this question, we introduced IN double mutants EH170,1AA, EK170,3AA, and HL171,2AA into an HIV-1 RT/IN trans-complemented single cycle replication system [292]. Briefly, each of these IN double mutants was first introduced into a CMV-Vpr-RT-IN expression plasmid. The VSV-G pseudotyped HIV-1 single cycle replicating viruses containing these individual IN double mutants and a luciferase gene, substituted for the Nef gene, were produced in 293T cells by co-transfecting each CMV-Vpr-RT-INwt/mut expression plasmid with RT/IN-deleted HIV provirus NLlucΔBgl/ΔRI, and a VSV-G expression plasmid. Then, the same amount of virus (normalized by p24 gag levels) was used to infect C8166 CD4+ T cells, and the level of infection was monitored by measuring the luciferase activity. The results showed that the mutant EK170,3AA, which can efficiently bind to both chromatin and LEDGF/p75 (Figure 13B and C), displayed about 30% replication capacity relative to the wild type virus (Figure 16A). The chromatin-bound

but LEDGF interaction defective IN mutant virus, EH170,1AA, induced a low level of infection, whereas the HL171,2AA mutant virus, which lost the ability to interact with both chromatin and LEDGF/p75, was non-infectious (Figure 16A). Moreover, real-time PCR analysis indicated that mutations introduced in the $_{170}$EHLK$_{173}$ did not significantly affect the reverse transcription step at 12 h post-infection (Figure 16B). These data suggest that while IN/LEDGF/p75 interaction is important for a productive HIV-1 replication, the IN-mediated LEDGF/p75-independent chromatin binding is still able to sustain a low level viral infection.

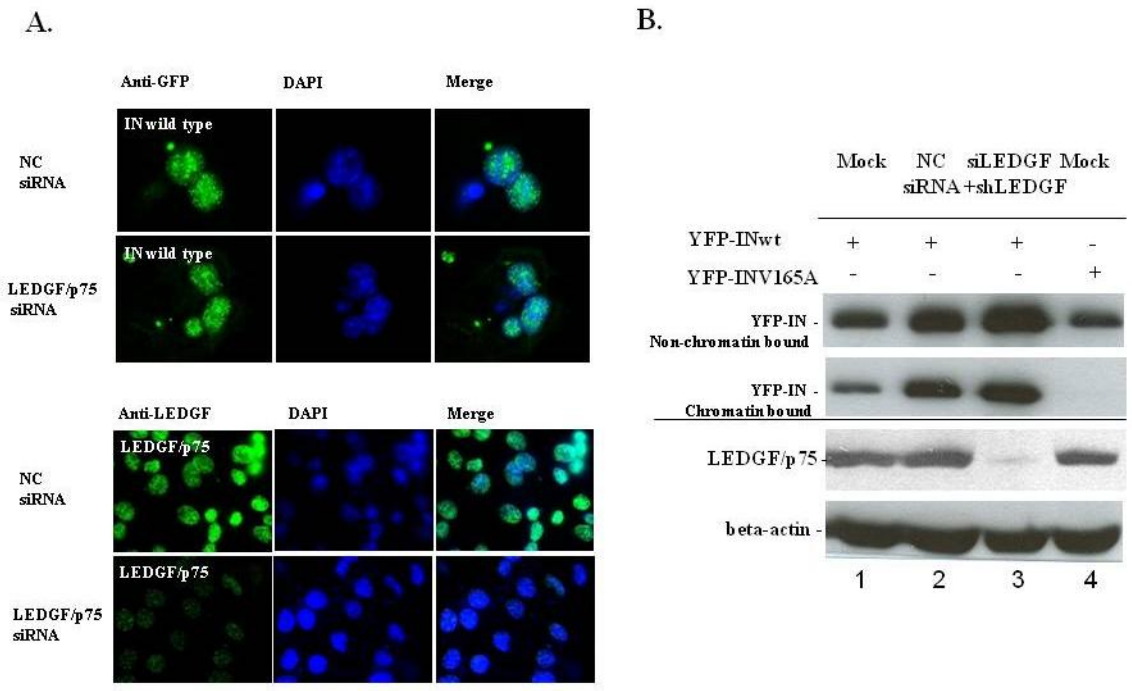

Figure 15 LEDGF/p75 is not required for chromatin binding of IN. A). Transient knockdown of LEDGF/p75 by siRNA had no effect on IN nuclear localization. 293T cells were transfected with either 20nM negative control (NC) siRNA or 20nM si-LEDGF PSIP1HSS146003 for 24 h before transfection with CMV-YFP-IN wild type. At 48 h post-transfection, cells were fixed, permeabilized and detected for YFP-IN and LEDGF/p75 expression by using anti-GFP or anti-LEDGF antibodies. The nuclei were stained with DAPI. B). Analysis of chromatin binding affinity of IN on LEDGF/p75 knockdown cells. The lentiviral shRNA-mediated LEDGF/p75 stable knockdown 293T cells were transfected with 20nM si-LEDGF for 48 h and further transfected with YFP-IN wild type or mutant V165A and were analyzed for its chromatin binding affinity. In parallel, cells were either mock-transfected or transfected with negative control siRNA to study chromatin binding of YFP-IN wild type. The chromatin bound and non-chromatin-bound fractions of YFP-IN wild type or V165A were showed as indicated. The

LEDGF/p75 expression level in each sample was verified by WB with anti-LEDGF antibody. Endogenous beta-actin was used for normalization of sample loading.

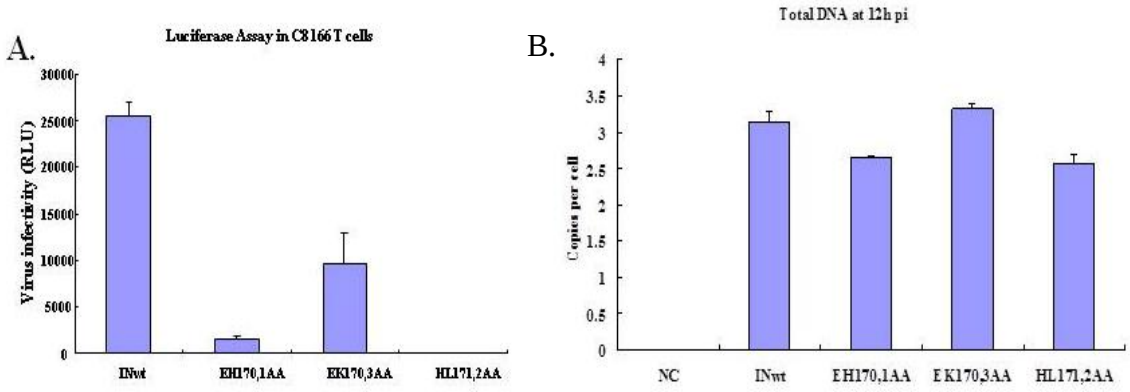

Figure 16 A) The differential replication profiles of IN mutant viruses within the loop $_{170}$EHLK$_{173}$ on HIV-1 single-cycle replication. To test the effect of different IN mutant viruses on HIV-1 infection, C8166 T cells were infected with equal amount of VSV-G pseudotyped single round replication virus containing various IN mutant (adjust by p24 level) for 48 h. 1x10^6 cells were collected and cell-associated luciferase activity was measured by luciferase assay at 48 h post-infection . Data represent mean ± standard deviation (SD) (n = 3). B) Followed by 12 h infection with single cycle replicating viruses on dividing C8166 T cells, total DNA was extracted and amplified for total viral DNA and human β-globin gene using corresponding primers by real-time PCR. Total HIV-1 DNA levels were expressed as copy numbers per cell, with DNA template normalized by the amplification of the β-globin gene. NC: negative control or 70 °C inactivated wt virus.

5.5 Discussion

In the present study, we investigated the interactions of various IN mutants with host cell chromatin and LEDGF/p75 by cell-based chromatin binding and Co-IP assays. In addition to previously described LEDGF/p75-binding defective IN mutants V165A, A179P, KR186,7AA [169, 172, 292], this study also identified several new IN mutants, including K159P, V176A and I203P, which reside in α4 to α6 helices of IN that lost the ability to bind to both chromatin and LEDGF/p75. Interestingly, we also found that several IN mutations, H171A, L172A and EH170,1AA, within the loop region $_{170}$EHLK$_{173}$ of IN, impaired the interaction with LEDGF/p75, but retained chromatin binding ability. This suggests that the IN is able to bind chromatin independently of LEDGF/p75. Consistently, our combined knockdown approach for LEDGF/p75 also failed to dissociate IN from chromatin. Moreover, we have also tested the effect of these IN mutants on HIV-1 infection, and our results revealed that the viruses harboring the IN mutants incapable of binding chromatin completely lost infectivity. However, viruses bearing IN mutants with chromatin-binding ability still sustained low levels of viral infection. All of these results clearly indicated that while the LEDGF/p75-binding ability of IN is important for productive HIV-1 replication, the IN has the ability to bind chromatin in a LEDGF/p75-independent manner and is sufficient to sustain a low level of HIV-1 infection. Taken together, these data confirmed the hypothesis that LEDGF/p75 is not strictly required for IN binding to chromatin.

5.5.1 Mutational analysis of chromatin- and LEDGF/p75-binding of IN within CCD

The results showed that IN mutants K159P, A179P and I203P located at the α-helices 4, 5 and 6 specifically affected both chromatin- and LEDGF/p75-binding abilities. Since introducing Pro often bends the amino-acid backbone and affects the secondary structure of the protein, it could be possible that introducing Pro mutations disrupts α-helix formation and hampers both chromatin- and LEDGF/p75-binding abilities. Indeed, this could be the case for the IN mutant I203P because another mutant I203A was able to efficiently bind host DNA and LEDGF/p75 (Figure 11 and 12 compare lane 14 to 15). However, given the fact that both A179P and A179I lost binding to host chromatin, the A179 residue may be directly involved in interacting with host chromatin (Figure 11 and data not shown for A179I). Nevertheless, the chromatin-binding phenotype of K159P, A179P and I203P IN mutants suggest the involvement of α-helices 4, 5 and 6 of IN in host DNA recognition. Two other IN mutants that need to be addressed are KR186,7AA and F185A. We have previously shown that the IN mutant KR186,7AA was severely impaired in both chromatin- and LEDGF-binding affinities [292]. In this study, we identified another mutant F185A that displayed a significant reduction in the interaction with LEDGF and chromatin, but to a lesser extent than that of KR186,7AA. The K186 and R187 of IN, by crystallographic studies, are known to lie in the dimer-dimer interface of IN [98, 99] and F185 has been implicated for tetramerization of IN [320]. So, mutations at F185, K186 and R187 might affect IN oligomerization and further impair its chromatin binding affinity. In addition, a recent study by Merad et al. revealed that a helix-turn-helix (HTH) (residues 149-186) motif consists of two helices (helix 4 and helix 5) and that the loop in between is involved in recognition of viral DNA [312]. Interestingly, in our study, IN mutants K159A, V165A, V176A, A179P, KR186,7AA are

located within this region and were identified as chromatin-binding defective mutants. Thus, the chromosomal attachment site within the IN CCD may also center on IN α-helix 4 to α-helix 5, and this HTH motif could be critical for the recognition of both viral and host DNA. However, how IN recognizes and binds both viral and host DNA sequence to form an active integration complex remains an open question and requires more detailed computational, experimental and structural investigations.

5.5.2 Uncoupled chromatin- and LEDGF/p75-binding of IN

However, the functional roles of LEDGF/p75 and its potential correlation with chromatin binding of IN are of interest in our present study. It is well established that LEDGF/p75 serves as an IN-to-chromatin tethering factor, driving PICs to transcriptionally active regions of host chromosomes [147, 168, 170]. Our previous results showed that chromatin binding defective IN mutants (V165A, A179P, KR186,7AA) also fail to interact with LEDGF/p75, suggesting that LEDGF-binding of IN might be linked to the chromatin-binding affinity of IN [292]. Here, we attempted to select more IN mutants to map both chromatin- and LEDGF/p75-binding sites within the CCD of IN. Results showed that most of the IN mutants tested in this study lost both chromatin-binding and LEDGF/p75-interacting abilities, highlighting the importance of LEDGF/p75 as a tethering factor for IN chromatin targeting. Interestingly, two IN mutants, H171A and L172A within the CCD of IN, displayed a different phenotype; they could not efficiently interact with LEDGF/p75 yet still could bind chromatin (Figure 11 and 12, see also Table 1). This raises the possibility that the HIV-1 IN may still be able to target chromatin in the absence of LEDGF/p75 association. Because H171 and L172 are located within or

close to the loci of IN/LEDGF interface (α4/5 connector residues 166–171) [171], we next focused on detailed chromatin- and LEDGF-binding affinities within the IN region $_{170}$EHLK$_{173.}$ For this purpose, four IN double mutants, EH170,1AA, EK170,3AA, HL171,2AA and HK171,3AA, were tested. Indeed, it was shown again that the IN mutant EH170,1AA showed relatively high affinity with host chromatin but was unable to bind LEDGF/p75 effectively, while the IN mutant HL171,2AA had defects on both the chromatin- and LEDGF-binding affinities (Figure 13). These results suggest that the HIV-1 IN is able to bind chromatin independently of LEDGF/p75.

Because the IN mutants H171A, L172A and EH170,1AA bound to chromatin but not LEDGF/p75, we further reconfirmed the LEDGF/p75-independent chromatin binding of wild type IN using the LEDGF/p75- knockdown cells. Our results showed that the efficient knockdown of LEDGF/p75 had no significant effect on IN-chromatin association, suggesting that the chromatin binding of IN might still take place in the absence of LEDGF/p75. Meanwhile, we ruled out the possible effect of LEDGF/p75 knockdown on nuclear translocation of HIV-1 with wild type IN by observing the intracellular localization of all the IN fusion proteins using immunostaining, which is indeed consistent with the previous observation [154]. Most likely, the IN is still able to target chromatin without preferential targeting sites in the absence of LEDGF/p75. Consistently, previous studies have highlighted that the role of LEDGF/p75 during HIV-1 integration is advantageous to HIV-1 integration but could be nonessential to the process of integration [120, 170]. We speculate that, without the LEDGF/p75 tethering, IN might still be able to bind chromatin, but it might lack the preferential selection site. Also, it is possible that other unknown cellular factor(s) might contribute to the

chromatin targeting of IN; such proteins should harbor both DNA-binding and IN-binding domains similar to that of LEDGF/p75. Further efforts are underway to seek new cellular partners involved in IN-to-chromatin targeting.

5.5.3 A strict correlation between chromatin binding of IN and HIV-1 fitness

In an attempt to correlate IN chromatin-binding ability to its effect on virus infection, we introduced IN mutants EH170,1AA, EK170,3AA or HL171,2AA into a VSV-G pseudotyped HIV-1 single cycle replicating virus and investigated their effects on HIV-1 infection. As expected, viruses containing the IN HL171,2AA mutation, which lost both LEDGF/p75- and chromatin-binding abilities, are unable to replicate (Figure 16). This result is consistent with previous reports in which the impaired integration of proviral DNA into host cell chromatin accounted for the replication defect of the L172 mutant virus [321, 322]. Interestingly, another batch of viruses harboring the IN EH170,1AA mutation, which fail to associate with LEDGF/p75 but are still able to interact with chromatin, retain the infectivity towards the susceptible cell lines, although at a low efficiency. This suggests that the chromatin association of IN, rather than LEDGF/p75 binding, is essential for HIV-1 infection. These results are consistent with the previous study by Shun and his co-workers in which the LEDGF-null mouse embryo fibroblasts were able to support approximately 10% of HIV-1 integration compared to control cells [140]. These results again highlighted the importance of LEDGF/p75-binding property of IN during HIV-1 replication. It is possible that the LEDGF-independent chromatin binding of IN is still able to target viral PICs to host chromatin, but, without the escort of LEDGF/p75, such IN-mediated “nonspecific” chromatin binding is less efficient and/or could not efficiently target viral PICs to transcriptionally active sites in the chromatin

and mediate a productive viral replication. Another interesting question is how IN is still able to interact with host chromatin under a very low level of LEDGF/p75. Whether it is through IN directly binding to host DNA or whether it requires other undefined cofactor(s) for this process remains unclear and requires more detailed study. Successful elucidation of the mechanism underlying how HIV-1 IN possesses a LEDGF/p75-independent chromatin binding and identification of other IN-interacting cofactors involved in this process will contribute to a better understanding of the action of IN during HIV-1 replication and aid in development of efficient and comprehensive anti-HIV strategies.

Chapter 6 (Published the paper entitled "Host protein Ku70 binds and protects HIV-1 integrase from proteasomal degradation and is required for HIV replication" in J Biol Chem. 2011 May 20;286(20):17722-35.)

Increased stability of HIV-1 integrase by host protein Ku70 against proteasomal degradation and its impacts on HIV replication

6.1 Rational

Recently, considerable interest has been focused on the functional interaction between IN and host cellular proteins in the hope of disrupting their interactions, thereby blocking HIV-1 replication. In an attempt to identify host cellular partners for IN, several research groups have identified a number of IN cofactors using the yeast two-hybrid system, Co-IP assays, or *in vitro* reconstitution of the enzymatic activity of salt-stripped PICs [150-156]. A recent study by Studamire et al. found that 12 cellular proteins, including Ku70, could bind to the INs of both the M-MLV and HIV-1 through screening with a yeast two-hybrid system [150]. However, whether these cellular cofactors are associated with HIV-1 IN during HIV replication and their functional relevance remain unknown.

Ku70 is well known as a DNA repair protein and part of the NHEJ pathway. For most biological functions in which Ku70 participates, Ku functions as a heterodimer consisting of Ku70 and Ku80, named according to their respective molecular weights of 70 and 80 kD. As part of NHEJ pathway, Ku70 has been suggested to participate in the gap repair between viral DNA and host genome during HIV-1 integration [205]. Studies have shown that the NHEJ pathway is important for retroviral transduction or infection

and for the cell survival of HIV-1 infected or transduced cells [204, 206-209]. For example, HIV-1-based vector transduction or infection was markedly reduced in cells deficient in Ku80, DNA-PKcs, Xrcc4 or ligase IV [206, 208]. Moreover, NHEJ activity is required for 2-LTR circle formation, and Ku70 has been detected in M-MLV PICs [19, 208, 210, 211]. Ku80 was also shown to suppress HIV transcription by specifically binding to a negative regulatory element within the LTR [212]. All of these observations suggest that Ku70 or the K70/80 heterodimer may be involved in HIV-1 infection by affecting multiple steps of the viral replication cycle, such as integration. In addition, a novel deubiquitinating (DUB) enzymatic activity of Ku70 was recently described, in which Ku70 has a regulatory effect on Bax-mediated apoptosis by decreasing the ubiquitination of Bax and blocking Bax from proteasomal degradation [323]. However, whether Ku70 also exerts a DUB effect on other identified binding partners of Ku70 and how Ku70 interacts with the ubiquitin-proteasome pathway to deubiquitinate protein substrates are still unclear.

6.2 Hypothesis

Two hypotheses of this project are: 1) host protein Ku70 interacts with HIV-1 IN and 2) Ku70 is able to regulate IN expression and impacts on HIV-1 replication.

6.3 Objectives

1 Investigate whether host protein Ku70 regulates HIV-1 IN expression.

2 Verify the interaction of Ku70 and IN in transfected overexpression system and under physiological infection condition.

3 Examine the roles of Ku70 in HIV-1 replication.

6.4 Results

6.4.1 Cellular protein Ku70 protects HIV-1 IN from proteasomal degradation

As a part of the NHEJ machinery, the host protein Ku70 has been shown to participate in HIV integration and in the circularization of unintegrated viral DNAs [209, 211]. Interestingly, based on the results of a yeast two-hybrid assay, a recent study indicated that HIV-1 IN may bind to Ku70 [150], suggesting a direct association between HIV-1 IN and Ku70. To further investigate this viral/host protein interaction, we coexpressed Ku70 (T7-tagged Ku70) and HIV-1 IN (IN-YFP) in 293T cells and analyzed their interaction after 48 h of transfection. Surprisingly, our results revealed that T7-Ku70 overexpression significantly increased IN expression (Figure 17A, lanes 1 and 2). However, the coexpression of Ku70 with another HIV-1 protein, MA (MA-YFP), did not change the MA expression level (Figure 17A lanes 3 and 4). This suggests that Ku70 is able to increase IN expression. Alternatively, Ku70 could protect the IN protein from degradation [216].

To further test whether endogenous Ku70 could exert the same activity and whether it is due to a protective effect, we first knockdown the Ku70 expression using specific siRNA in 293T (Figure 17B) or HeLa cells (Figure 17C) and checked the level of GFP-IN expression by WB or fluorescence microscopy (Figure 17B and 17C). To increase IN expression under normal conditions, we used a pAcGFP-IN with a codon-optimized IN sequence (GFP-INopt). The results showed that IN expression in Ku70- knockdown cells was significantly decreased when compared with IN expression in siNC cells (Figure 17B, compare lanes 1 and 2; 1C, compare A1–3 and B1–3). Interestingly, in the

presence of the specific proteasome inhibitor MG-132, IN expression in Ku70-knockdown cells was remarkably increased, reaching levels similar to those in siNC-transfected 293T and HeLa cells (Figure 17B, compare lanes 4 and 3; 1C, compare C1–3 and D1–3). Thus, these results clearly indicate that Ku70 is able to protect IN from proteasomal degradation.

6.4.2 Ku70 is able to interact with HIV IN in both 293T cells and HIV-1 infected T cells

Given that Ku70 is able to protect HIV-1 IN from proteasomal degradation, we next tried to reveal the molecular mechanisms underlying this effect. Because Ku70 has been implicated as an HIV-1 IN cofactor [150], it is possible that IN could escape from the host proteasomal degradation machinery by directly interacting with Ku70. Therefore, we further investigated the interaction between HIV IN and host protein Ku70 under more physiological conditions by using a Co-IP approach in 293T cells and HIV-1 infected CD4+ C8166 T cells.

First, a CMV-Ku70 expressor and GFP or GFP-INwt plasmid were co-transfected into 293T cells. To prevent IN degradation, MG-132 (10μM) was added to the cells at 12 h before cell lysis. The cells were then lysed and immunoprecipitated with anti-GFP followed by a WB using anti-Ku70. We found that GFP-IN, but not GFP, was able to pull down Ku70 (Figure 18A). Given that both Ku70 and IN are DNA binding proteins, we added DNase I in the cell lysate during the Co-IP assay and found that GFP-IN still bound to Ku70 with DNase I treatment, suggesting a direct protein-protein interaction (data not shown). Considering that the protein-overexpression system might not

necessarily reflect normal functional binding of IN/Ku70, we also tested the authentic interaction of IN/Ku70 in HIV-1-infected cells (Figure 18B). To do this, C8166 CD4+ T cells were infected with HIV-1 HxBru or HxBru-IN-HA viruses. In the provirus HxBru-IN-HA, an HA tag was inserted at the C-terminus of IN [109], allowing us to pull down HIV-1 IN-associated cellular proteins by using anti-HA antibody in the Co-IP assay. At 72–96 h post-infection, the C8166T cells were lysed, and the cell lysates were subjected to Co-IP assay to detect IN-bound endogenous Ku70 in the infected cells. The results showed that Ku70 was coprecipitated with IN-HA from HxBru-IN-HA-infected cells, but not from mock C8166T cells or HIV HxBru-infected C8166 T cells (Figure 18B, upper panel). The same immunoblot was reprobed with anti-HA to detect the presence of IN-HA (Figure 18B, second panel). Similar levels of endogenous Ku70 and viral Gag-p24 in HxBru- and HxBru-IN-HA-infected cells (Figure 18B, third and fourth panel) served as input controls. Together, these results convinced us that the DNA repair protein Ku70 is an authentic, newly described host cofactor for IN.

To delineate the Ku70-binding domain of IN, seven previously described IN N-terminal or C-terminal deletion mutants, including GFP-IN 1–230, 1–250, 1–270, 50–288, 112–288 and one substitution mutant, GFP-INKR186,7AA [285, 292], were used to test Ku70 binding ability (Figure 18C). The results revealed that GFP and GFP-IN1-230 did not bind to Ku70 (Figure 18D, lanes 1, 3, and 8), while other truncated GFP-IN mutants (1–250, 1–270, 50–288, and 112–288) and the point mutant KR186,7AA still retained their Ku70-binding ability (Figure 18D, lane 2, 4–7, and 9). KR186,7AA, a well-characterized oligomerization-defective mutant of IN [79, 89], still bound Ku70,

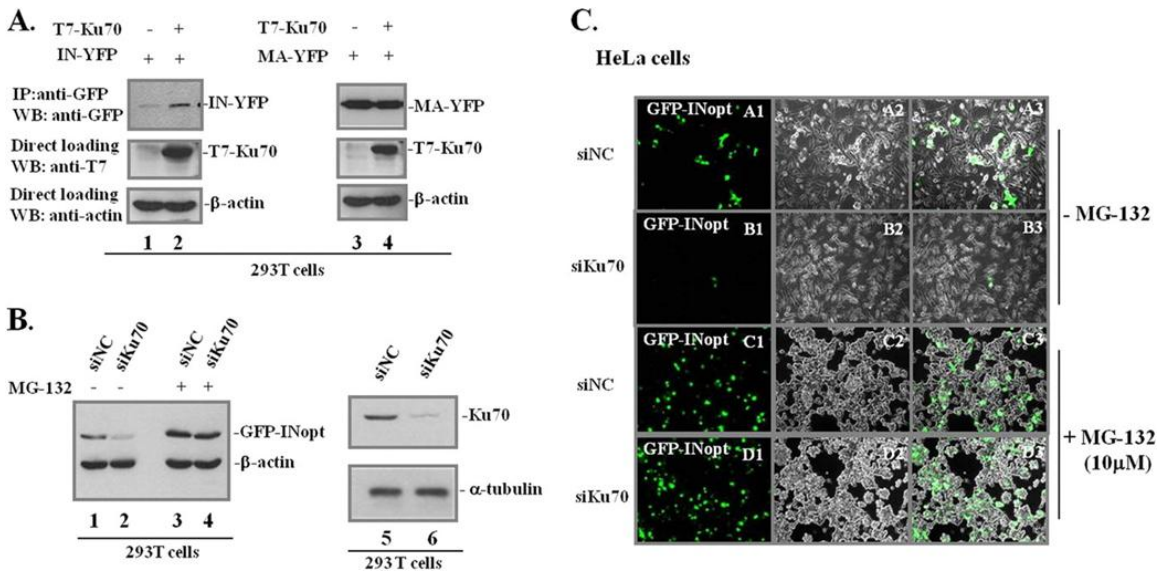

Figure 17 Effects of Ku70 on the stability of HIV-1 IN. A) 293T cells were cotransfected with IN-YFP or MA-YFP and T7-Ku70 wt for 48 h. Expression of IN-YFP (lanes 1 and 2) and MA-YFP (lanes 3 and 4) were detected by IP with rabbit anti-GFP antibody and anti-GFP-HRP antibody in a WB (upper panel). An aliquot of cells (about 5%) was collected and checked for T7-Ku70 wt and β-actin expression (middle and lower panels). B) Proteasome inhibitor MG-132 treatment restores IN expression in Ku70- knockdown 293 T cells. After transfection with 5 nM siKu70 or the siRNA negative control (siNC) for 48 h, 293T cells were transfected with optimized IN (GFP-INopt) for another 48 h. The control and knockdown cells were treated with or without 1 μM MG-132 for 12 h prior to harvesting. The expression of GFP-INopt was detected by direct loading of protein samples onto a 10 % SDS-PAGE gel using HRP-conjugated anti-GFP antibody in a WB (lanes 1–4), and the same membrane was probed with anti-β-actin antibody to assess protein loading. Ku70 knockdown efficiency was detected by mouse anti-Ku70 antibody in WB (lanes 5 and 6) and α-tubulin was used as the protein-loading control in each sample. (C) The effects of Ku70 on HIV-1 IN expression were

confirmed in HeLa cells by direct fluorescence. HeLa cells were treated with 5 nM siKu70 or siNC for 48 h prior to transfection with GFP-INopt. After another 48 h, GFP-positive HeLa cells without MG-132 (A1-4, B1-4) or with 12 h treatment of 10 μM MG-132 (C1-4, D1-4) were examined by fluorescence microscopy.

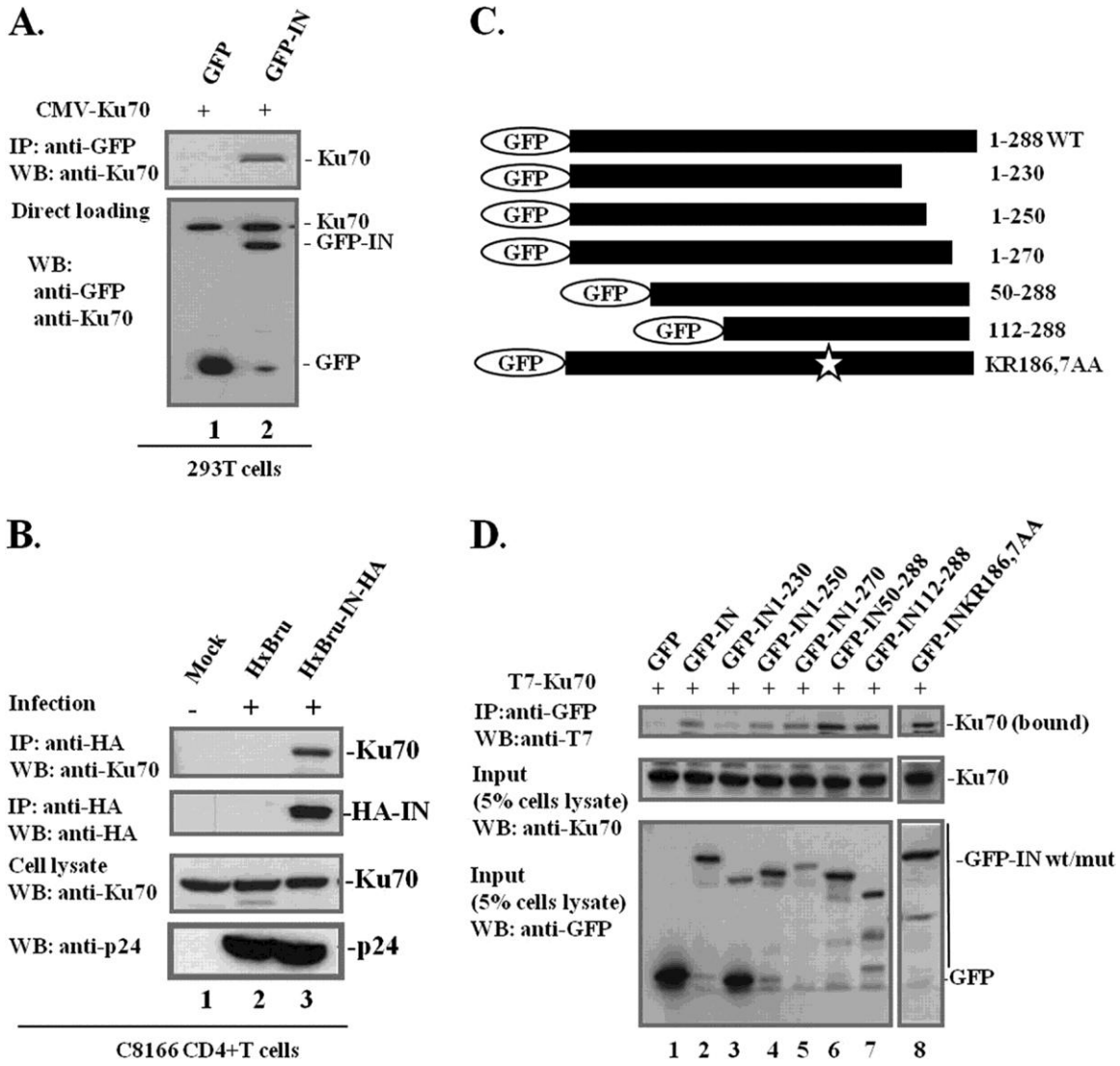

Figure 18 IN interacts with Ku70 in mammalian cell lines and in HIV-1-infected CD4+ T-lymphocytes, and the interaction is through the C-terminus of IN. A) Interaction of Ku70 with GFP-IN in 293T cells. GFP or GFP-IN was coexpressed with CMV-Ku70 as indicated into 293T cells for 48 h, and cells were treated with the proteasome inhibitor MG-132 (10 μM) for 12 h prior to Co-IP analysis. Co-IP was performed using rabbit anti-GFP antibody for immunoprecipitation and WB using anti-Ku70 to detect IN-bound Ku70 (Upper panel). GFP-IN and untagged pCMV-Ku70 were detected with the corresponding antibodies to assess the expression levels of IN and Ku70 (lower panels of

lanes 1 and 2). B) C8166 CD4+ T cells were mock infected or infected with HIV-1 HxBru or HxBru-IN-HA, as indicated. Cells were collected at 72 h post-infection, and the cell lysates were immunoprecipitated with anti-HA antibody and immunoblotted with anti-Ku70 antibody to detect IN-associated Ku70 in HIV-1-infected T cells (upper panel). The same membrane was then reprobed with anti-HA antibody to detect IN-HA expression (second panel). The same amount of cells was assessed for Ku70 expression prior to Co-IP, serving as the immunoprecipitation input (third panel). The p24 levels in the infected cells were checked with anti-p24 antibody (lower panel). C) Schematic representation of the GFP-IN wild type, seven IN deletion mutants and site mutant KR186,7AA. D) The C-terminus of IN is required for IN/Ku70 interaction. GFP or various GFP-IN wt/mutants were co-transfected with T7-Ku70 (lanes 1–8) into 293T cells, and their interactions were studied by Co-IP assay. Cells were treated with MG-132 at a concentration of 10 μM for 12 h prior to Co-IP analysis. IN-bound Ku70 was detected by IP with anti-GFP antibody and WB with anti-Ku70 antibody (upper panel). Total cell lysates were analyzed for GFP-INwt/mutants and Ku70 expression (middle and lower panel; input: 5% of total cell extract).

suggesting that multimerization of IN is not required for IN/Ku70 interaction (Figure 18D, lane 9). Overall, our analysis suggested that the C-terminal half of IN (IN112-288) is sufficient to bind to Ku70. Therefore, more detailed studies are still required to identify the critical motif(s) in HIV IN for Ku70 binding.

6.4.3 The Ku70 truncated mutant Ku701-430 interacts with HIV IN, but cannot form a heterodimer with Ku80

To define the IN-binding region within Ku70, expressors for four T7-tagged Ku70 deletion mutants (1–263, 1–430, 226–609 and 430–609) were constructed (Figure 19A) and cotransfected with GFP-IN or the expressor into 293T cells. Based on a previous mutational analysis indicating that the minimum region for DNA binding within the Ku70 core region was estimated to be in the around aa 263–430, 1–263 and 1–430 mutants were constructed; Ku70 226-609 was shown to be defective for DNA-PK activity and DNA binding in the same study [324]. The Ku70 430-609 mutant was sufficient for the heterodimerization of Ku70/80 but lost DNA end-binding ability in the two-hybrid analysis [182]. In parallel, 293T cells cotransfected with T7-Ku70 and GFP plasmids were used as a negative control. The results showed that T7-Ku70wt and deletion mutant T7-Ku701-430 were coimmunoprecipitated with GFP-IN (Figure 19B, upper panel). Interestingly, T7-Ku70 1-430 displayed a higher binding affinity for IN than T7-Ku70wt (Figure 19B, upper panel, compare lane 4 to 2). Because the N-terminal truncation (1–263) of Ku70 did not interact with IN (Figure 19B, upper panel, lane 3), whereas T7-Ku70 1-430 showed a strong binding affinity, the aa 263–430 region of Ku70 is probably necessary but not sufficient for IN interaction. However, another

Ku70 mutant, 226–609, which also encodes the aa 263–430 region, failed to interact with GFP-IN (Figure 19B, upper panel, lane 5), suggesting that another important binding domain might exist within the N-terminus (1–226) of Ku70. Therefore, both the N-terminal domain and the core domain of Ku70 are suggestive of binding surface for IN.

Ku70 forms a heterodimer with Ku80, the heterodimerization which has been shown to contribute to many cellular processes. For example, the heterodimerization of Ku70/80 is essential for activating DNA-PK and DNA repair and important for their nuclear translocation [324, 325]. One might ask whether Ku80 is also involved in the IN/Ku70 interaction or if IN interacts with Ku70 indirectly through Ku70/80 heterodimerization. To address this question, we performed a Co-IP assay in which T7-Ku70wt or T7-Ku701-430 was transfected into 293T cells. At 48 h post-transfection, cell lysates were immunoprecipitated with anti-T7 antibody followed by WB with an anti-Ku80 antibody. In agreement with our previous findings, T7-Ku70 wt was able to pull down endogenous Ku80 (Figure 19C, lane 2, upper panel). Interestingly, the deletion mutant T7-Ku701-430, which efficiently binds IN, could not form a heterodimer with Ku80 (Figure 19C, lane 3, upper panel). This result indicates that T7-Ku701-430 binding to IN was independent of Ku80. Thus, the heterodimerization of Ku70/80 may not be required for IN/Ku70 interaction.

6.4.4 Ku 70 protects IN from degradation by reducing the total ubiquitination level in the host cells and by IN/Ku701-430 binding

The results above showed that Ku70 binds and protects IN from proteasomal degradation;

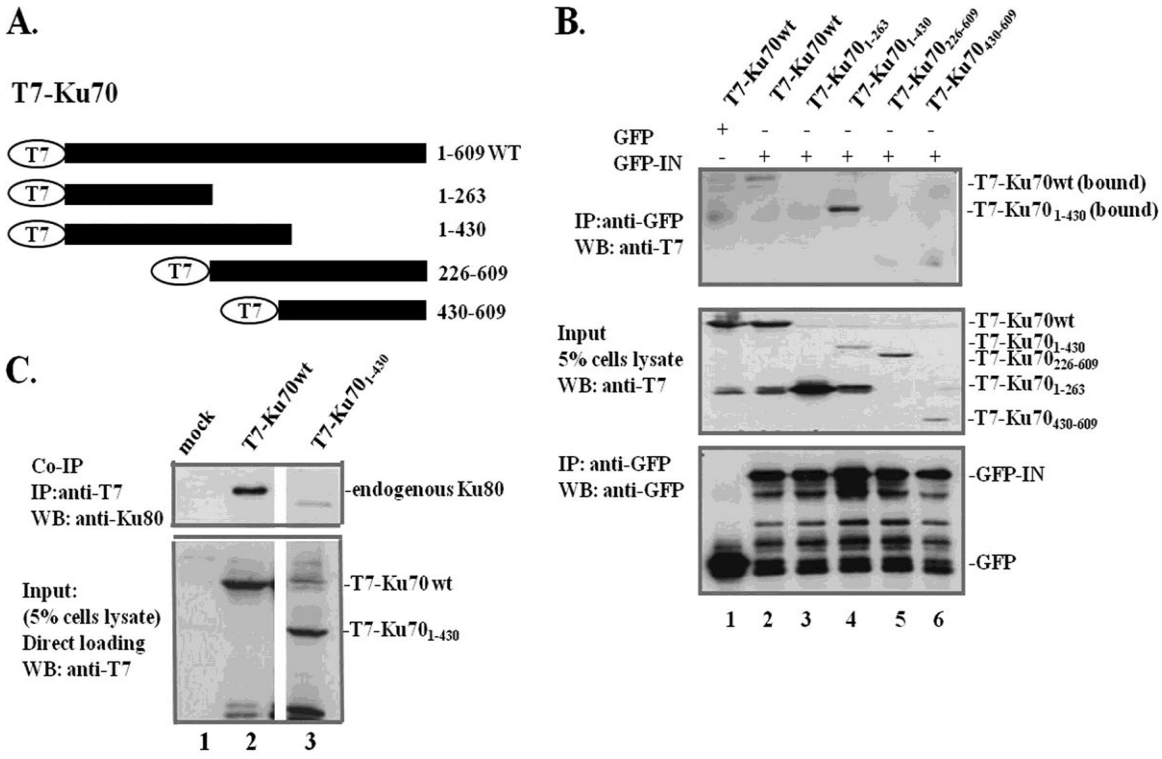

Figure 19 The N-terminus (1–430) of Ku70 binds IN and IN/Ku70 interaction is independent of the heterodimerization of Ku70/80. A) Schematic diagrams depict the different T7-Ku70 wt/mutant constructs used in the domain-mapping experiments. The full length of T7-Ku70 is shown at the top, as indicated. B) Interaction of GFP-IN with T7-Ku70 wt/mutants. GFP or GFP-INwt was cotransfected with T7-Ku70wt/mut in 293T cells for 48 h. MG-132 (10 μM) was added to the cells 12 h prior to cell lysis to enhance protein expression. The Co-IP assay was performed to map the IN binding region in Ku70 using anti-GFP antibody IP, and a WB using anti-T7 antibody was performed to detect IN-bound T7-Ku70wt/mut (upper panel). GFP, GFP-IN and T7-Ku70wt/mut expression were checked by immunoblotting with anti-GFP or anti-T7 antibodies, respectively (middle and lower panels). C) T7-Ku701-430 cannot form a heterodimer with Ku80. 293T cells were mock transfected or transfected with T7-Ku70 wt and the aa 1–430 truncation mutant for 48 h. Heterodimerization of Ku70/80 was

determined by Co-IP with anti-T7 antibody and WB using mouse anti-Ku80 antibody to detect T7-Ku70-bound endogenous Ku80 (upper panel). About 5% of the cell lysates (input) were checked for the expression of T7-Ku70wt/mutants by WB using anti-T7 antibody (lower panel).

however, the detailed mechanisms underlying the degradation of IN through the ubiquitination-proteasome pathway remain unclear. Proteins tagged with ubiquitin can be monoubiquitinated or polyubiquitinated. The polyubiquitination chains are formed between the C-terminal residue Gly 76 of ubiquitin and any other internal Lys within the ubiquitin molecule (Lys 6, 11, 27, 29, 33, 48 or 63) through an isopeptide bond [326, 327]. The two most important polyubiquitination chains are the K48- and K63-linked chains, with the K48-linked polyubiqutination chain recognized by the 26S proteasomal pathway for degradation and the K63-linked polyubiquitination chain implicated in postreplicative DNA repair [328]. To analyze the ubiquitin-proteasome pathway involved in IN degradation, HA-Ub mutants K48R and K63R were included in our study, having mutations of K48 and K63 to Arg that were expected to disrupt G76–K48 and G76–K63 polyubiquitination-chain formation, respectively. First, HA-Ubwt or mutant HA-UbK48R, HA-UbK63R were cotransfected with the IN expressor (GFP-IN) into 293T cells. At 48 h post-transfection, cells were lysed in 0.25% NP-40 and subjected to a Co-IP assay using anti-GFP antibody to pull down GFP-IN, followed by WB with anti-GFP and anti-HA antibodies to detect HA-Ub tagged IN-associated proteins (Figure 20A, upper and middle panels). Simultaneously, HA-Ub expression levels in the cells were also checked (Figure 20A, lower panel). The data showed that GFP-IN protein levels in the HA-UbK48R overexpression cells were much higher than in HA-Ubwt- and HA-UbK63R-transfected 293T cells (Figure 20A, upper panel; compare lane 3 to lanes 2 and 4). Similarly, the ubiquitination level of IN-associated protein expression was the highest in the HA-UbK48R-transfected sample (Figure 20A, middle panel). The HA-tagged ubiquitination signal (Figure 20A, middle panel) was from a pool of ubiquitinated

IN and unknown IN-bound cellular proteins. The observation of increased levels of ubiquitinated IN-associated proteins in HA- UbK48R cotransfected cells was expected given the fact that the immunoprecipitate input or GFP-IN (Figure 20A, upper panel) was the highest and that all the IN-associated proteins subjected to the UPS for degradation were accumulated due to the defective K48-linked polyubiquitination proteasome degradation pathway. However, similar levels of HA-Ubwt, HA-UbK48R, and HA-UbK63R were detected in the cell lysates (Figure 20A, lower panel). The highest IN expression was detected in the HA-UbK48R expression cells, clearly suggesting that GFP-IN is degraded though the K48-linked polyubiquitination proteasomal degradation pathway.

To further investigate how Ku70 affects IN stability, we studied the ubiquitination level of IN in the presence of both T7-Ku70 and HA-Ub. As T7-Ku701-430 showed a strong binding affinity with IN (Figure 19), we also included this T7-Ku70 deletion mutant to determine if the interaction of Ku70/IN plays a role in the stability and ubiquitination of IN. As expected, GFP-IN expression in the cells transfected with T7-Ku70wt and T7-Ku701-430 deletion mutants were 1.47±0.11 and 1.78±0.24 fold increased compared to cells transfected with the empty vector (Figure 20B, upper panel; compare lanes 3 and 4 to lane 2). The total ubiquitin expression level detected in the whole-cell extract was dramatically reduced in the presence of wild-type Ku70 (Figure 20B, lane 3, third panel). Consistently, the Co-IP data showed that ubiquitinated IN-bound proteins also remarkably reduced by 4.58±0.51 fold in the presence of wild-type Ku70 (Figure 20B, lane 3, second panel). Interestingly, T7-Ku701-430 was still able to protect IN (Figure 20B, first and third panels, lane 4) and significantly reduced the level of HA-tagged

ubiquitination signal in IN-bound proteins in the presence of T7-Ku701-430 by 1.18±0.19 fold (Figure 20B, second panel; compare lane 4 to lane 2; the percentage is normalized by total HA-ubiquitin level in the third panel), even though it did not affect the overall ubiquitin level in the cells (Figure 20B, third panel; compare lane 4 to lane 2). These results thus indicate that even though T7-Ku701-430 lacks the activity to reduce total ubiquitin level, as wild-type Ku70 does, it can protect IN by specifically reducing the ubiquitination of IN and its associated proteins. Based on these results, we conclude that Ku70 protects IN through two mechanisms: Ku70 reduces total ubiquitination level in the cells and reduces the ubiquitination of IN and IN-bound cellular proteins through their interaction.

We then tested whether Ku70 affects specific lysine-linked polyubiquitination proteasomal degradation pathways. The T7-Ku70 expressor was cotransfected with HA-Ubwt, HA-UbK48R, or HA-UbK63R into 293T cells. The results showed that, in the presence of Ku70, overall HA-Ub expression was still greatly reduced even though Ub cannot form K48- or K63-linked polyubiquitin chains (Figure 20C). This observation was confirmed by the fact that the down-regulation of Ku70 increased ubiquitin levels in the cells. We established a stable Ku70- knockdown CD4+ C8166 T-cell line by transducing C8166 CD4+ T cells with a lentiviral vector carrying Ku70 shRNA. In parallel, the control cell line was transduced with an empty lentiviral vector. After puromycin selection, control cells and Ku70- knockdown cells were checked for Ku70 knockdown efficiency by WB (Figure 20D, lower panel). These cells were then used to detect endogenous ubiquitin levels. The results showed that the ubiquitin levels were higher in the Ku70- knockdown cells than in the empty-vector-transduced cells (Figure

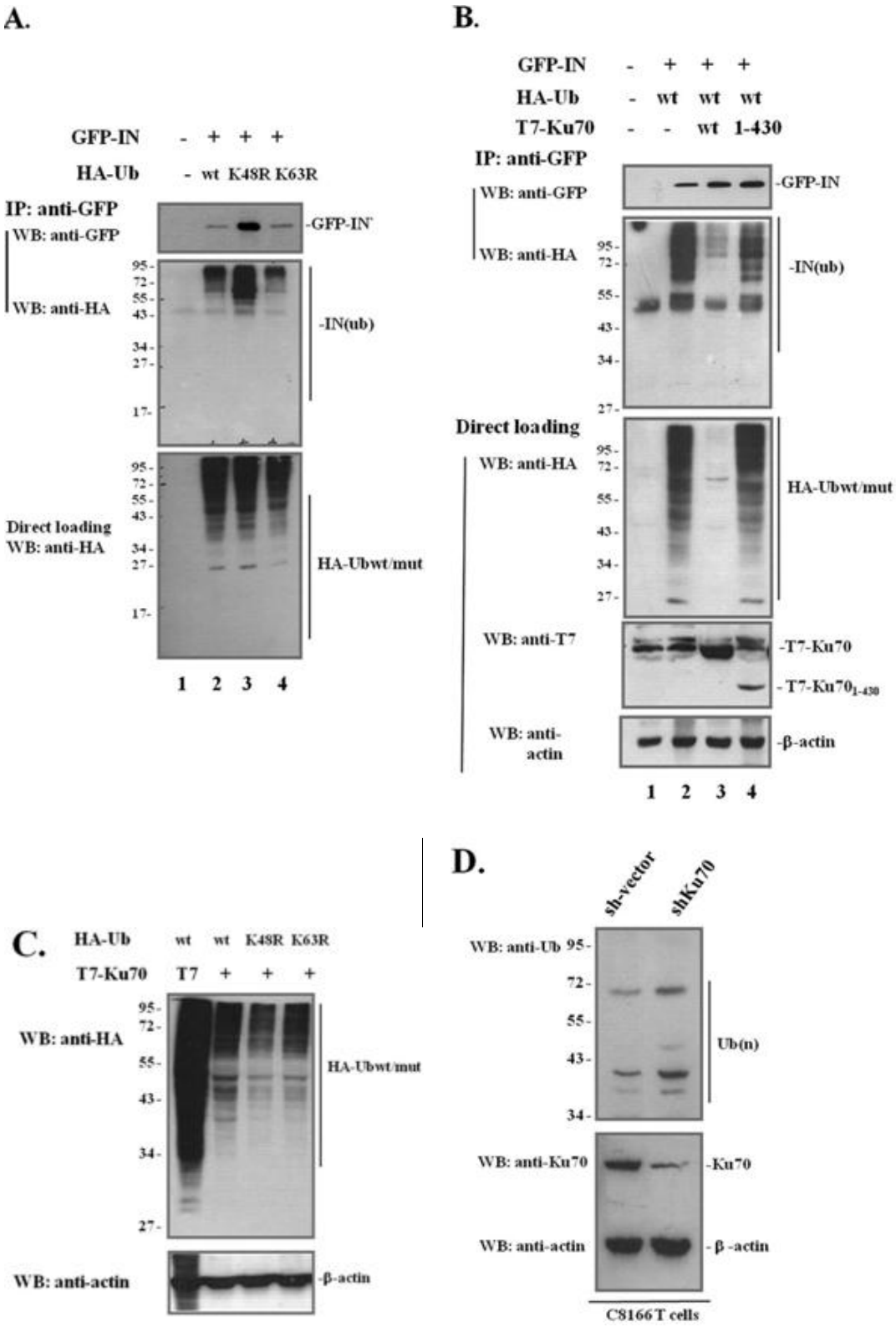

Figure 20 IN is degraded through the K48-linked polyubiquitination proteasomal

pathway, and Ku70 protects IN by reducing the overall ubiquitination level in the cells and partially blocking ubiquitination of IN and its bound cellular proteins. A) IN is degraded through K48-linked polyubiquitination. 293T cells were mock transfected or cotransfected with GFP-IN and HA-Ub wt or HA-UbK48R, HA-UbK63R for 48 h, as indicated. The ubiquitination levels of IN and IN-associated proteins were checked by IP with anti-GFP antibody and anti-HA antibody in a WB (middle panel). The same membrane was reprobed with anti-GFP antibody to detect GFP-IN expression in each sample (upper panel). About 5% of the cells prior to Co-IP were lysed in 0.5% NP-40, and the protein contents were subjected to WB to detect total HA-Ub levels in the cells using anti-HA antibody (lower panel). B) Ku70 protects IN from degradation by targeting the cellular ubiquitin-proteasome pathway and reducing ubiquitin binding to IN. 293T cells were mock transfected (lane 1) or cotransfected with GFP-IN, HA-Ubwt and T7 vector or T7-Ku70 wt, T7-Ku701-430 (lanes 2–4). The Co-IP assay was done at 48 h post-transfection using anti-GFP antibody to pull down IN and its associated proteins and using immunoblotting with anti-HA antibody to determine the ubiquitination level of IN and its associated proteins (second panel). The same membrane was reprobed with anti-GFP antibody to examine GFP-IN expression (first panel). Simultaneously, equal amounts of total cellular proteins (about 5% of the total cell lysates) were resolved on an SDS-PAGE gel and immunoblotted with anti-HA and anti-T7 antibodies to determine the expression levels of the transfected protein expressors (third and fourth panels). β-actin was used as a loading control (lowest panel). C) Reduction of ubiquitin level by Ku70 is independent of the K48- and K63-linked polyubiquitination proteasomal pathway. 293T cells were transfected with HA-Ubwt/mut with T7-Ku70 or T7 vector for 48 h. Cells

were lysed and analyzed for HA-Ub expression using anti-HA antibody in a WB. On the same membrane, β-actin was used as a protein-loading control. D) Endogenous ubiquitin levels were increased in Ku70-down-regulated cells. C8166 T cells were transduced with empty vector or lentiviral vector expressing a shRNA against human Ku70 and selected with 1 µg/mL puromycin. After one week of selection, equal amounts of control or stable Ku70- knockdown cell lines were collected and lysed. The expression levels of ubiquitin, Ku70 and beta-actin were assessed by WB using specific antibodies.

20D, upper panel). Taken together, we demonstrate here that Ku70 is able to protect HIV IN from degradation by down-regulating cellular ubiquitin levels and simultaneously preventing the ubiquitination of IN and its associated cellular proteins.

6.4.5 Ku70 knockdown impairs HIV-1 replication

Given that Ku70 is able to bind HIV IN and protects IN from degradation, it was interesting to test whether and how Ku70 contributes to HIV-1 replication. To do so, the empty-vector-transduced and Ku70-knockdown C8166 T cells were infected with pNL4.3-GFP+ viruses at an MOI of 0.5 (Figure 21A) or 0.05 (Figure 21B) for 2 h, and the viral replication kinetics were monitored. The infection was examined at different time intervals by harvesting virus-laden supernatant and checking for HIV-1 p24 antigen release. Figure 21A shows that at the MOI of 0.5, HIV infection in Ku70- knockdown C8166 T cells was reduced by approximately 50% at four and five days post-infection compared to the control cells (Figure 21A). However, when Ku70- knockdown cells were infected with a lower MOI of 0.05, viral infection was undetectable by measuring HIV p24 levels up to 11 days. However, in the control cells, viral replication peaked at day nine (Figure 21B). Interestingly, Ku70 knockdown completely inhibited low-MOI HIV-1 infection but only reduced high-MOI viral infection by 50%. This could be due to the fact that the infection of a large amount of viruses produced from normal T cells may at least partially overcome the shortage of Ku70 inside the target cells and establish an efficient first cycle of replication in Ku70- knockdown cells. If this is the case, we reasoned that virus produced from Ku70- knockdown cells should have a significantly lowered infectivity compared to the virus produced from the normal cells. To test this

possibility, we infected empty-vector-transduced and Ku70- knockdown C8166 T cells with the same amounts of pNL4.3-GFP+ virus (normalized by p24 ELISA) produced from either empty-vector or Ku70- knockdown C8166 T cells at an MOI of 5. At three days post-infection, virus-laden supernatants were harvested and checked for p24 antigen production by p24 ELISA. Consistent with the above results, when cells were infected with virus produced from empty-vector-transduced cells (empty-vector virus or E-virus), there was only a two fold difference in Ku70- knockdown cells compared to its infection in empty-vector cells (Figure 21C; compare bars 1 and 2). However, when empty-vector-transduced cells were infected with either E-virus or Sh-virus produced from Ku70- knockdown cells, there was an approximately five fold reduction of Sh-virus infection (Figure 21C; compare bars 1 and 3). Strikingly, Sh-virus infection in Ku70-knockdown cells exhibited more severe impairment, with a 16-fold reduction in viral infectivity compared with the E-virus infection in empty-vector-transduced cells (Figure 21C; Compare bars 1 and 4). Overall, this group of results indicates that down-regulation of Ku70 in both HIV-producing and target cells significantly impairs viral replication.

To investigate at which early step(s) of viral replication is blocked by Ku70- knockdown, infected C8166 T cells as described above (at an MOI of 5) (Figure 21C) were harvested at 24 h post-infection and DNA were isolated and assessed for late RT, 2-LTR circles and integrated DNA by qPCR [285]. Results revealed that Late RT products in the Ku70-knockdown C8166T cells did not showed significant reduction when compared with normal empty vector transduced cells ($p>0.05$, compare Figure 21D left, bar 3 to bar 2), while Late RT product was about 50% reduced compared with Ku70- knockdown cells

infected with sh-virus ($p<0.05$, compare Figure 21D left panel, bar 4 to bar 2). Strikingly, 2-LTR and integrated DNA in shKu70 knockdown cells infected with either normal E-virus or sh-virus were undetectable under current assay condition ($p<0.01$, Figure 21D middle and right panel, bar 3 and 4). 2-LTR circles formation is indicative of nuclear import [21], but it also requires proper circularization by host DNA repair enzymes [211, 329]. For example, depletion of NHEJ pathway components (Ku80, XRCC4, Ligase IV) resulted in 2-LTR formation undetectable or reduced [211, 329]. Undetectable 2-LTR circle formation in Ku70- knockdown cells infection with both normal E-virus and sh-virus could be due to insufficient DNA circularization by Ku70- knockdown in the target cells (Figure 21D, middle panel, bar 3 and 4). All of these results indicate that the presence of Ku70 is required for an efficient viral reverse transcription and necessary for viral integration.

6.4.6 Host protein Ku70 is incorporated into viral particles and stabilize IN expression

Because virus produced from Ku70- knockdown cells cannot infect C8166 cells efficiently, we first checked whether Ku70 knockdown leads to a defective maturation of the HIV-1 particle. The progeny viruses produced from empty-vector-transduced cells and Ku70- knockdown C8166 T cells were normalized by p24 value and loaded onto an SDS-PAGE gel. Simultaneously, infected C8166 T cells were lysed and also subjected to the same analysis. Next, anti-p24 and anti-IN antibodies were used to check for the presence of p24 and IN in the virions and infected cells. We did not detect any difference in the p24/IN ratio in the E-virus and sh-virus nor in the infected cells (Figure 22A,

upper panel). This suggests that the virus Gag-pol processing remained unaffected in the progeny virus produced from Ku70- knockdown cells. To examine the RT activity in the virions produced in Ku70- knockdown cells, the same amounts of viruses (normalized by p24 ELISA) from either empty-vector-transduced cells or Ku70- knockdown cells were lysed and subjected to a reverse-transcriptase assay (Roche). There was no significant difference in viral RT activity observed between normal and Ku70- knockdown cells (Figure 22A, lower panel). All of these data suggest that Ku70- knockdown does not affect the processing of Gag/Gag-Pol or virus maturation.

The observation that the Ku70- knockdown phenotype resulted in defective progeny virus (Figure 21) implies that Ku70 might be packaged into the progeny virus particles affecting HIV-1 replication. To test this hypothesis, we checked for the presence of Ku70 in the HIV virion. Briefly, the C8166 T cells were mock infected or infected with HIV pNL4.3-GFP at an MOI of 1 for 2 h. Virions were isolated four days after infection on a 20% sucrose gradient by ultracentrifugation (35,000 rpm) for 1.5 h. Virions were lysed in RIPA buffer and precipitated in 20% TCA to concentrate their protein contents before WB analysis. Simultaneously, the infected or uninfected C8166 T cells were lysed and analyzed by WB. Note that Ku70 was detected in the virions prepared from infected cells but not in the mock-infected cells (Figure 22B, upper panel). Additionally, the presence of p24 in the virions and the cells was detected by immunoblotting with anti-p24 antibody on the same membrane (Figure 22B, upper panel).

To verify that Ku70 incorporation into HIV-1 virion is mediated by IN, the VSV-G pseudotyped HIV-1 single cycle infection system was used. Vpr-RT-IN or Vpr-RT was transcomplemented with VSV-G and RT/IN/Env gene-deleted NLlucΔBglΔRI provirus

into 293T cells to produce IN+ and IN- virus (Figure 22C). The viruses were collected and ultracentrifuged through 20% sucrose at 35,000 rpm for 2 h. Since subtilisin treatment can effectively remove the proteins (either microvesicles or exosomes) outside the virions of purified HIV-1 virions [330], the IN+ and IN- virus were treated with 0.1 mg/ml subtilisin to remove potential contamination outside the virion. As shown in Figure 22B (lower figure), Ku70 expression is dramatically higher in IN+ virus than IN- virus without subtilisin treatment (compare lane 2 with lane 3, upper panel). However, when virions were treated with subtilisin, Ku70 was evidently detected in IN+ virus but not IN- virus (lane 5-6, upper panel). IN and p24 expression in both viruses were monitored by blotting with anti-IN and anti-p24 antibodies. Meanwhile, Ku70 expressions in mock transfected or transfected cells were assessed (Figure 22B. lower panel). Taken together, these results suggest that mediated by IN, Ku70 is incorporated into the HIV-1 particle.

The above results (Figure 22B and Figure 17) suggest that virus-associated Ku70 might stabilize IN or protect it from degradation in the infected cells. To investigate the protective roles of Ku70 on IN during viral infection, we assessed IN expression in the presence or absence of Ku70 during VSV-G pseudotyped HIV-1 single cycle infection. To do so, we have constructed Vpr-RT-IN-ProLabel (Vpr-RT-IN-PL) plasmid with ProLabel tag in the C-terminal of IN which enables us to quantify IN expression by measuring ProLabel activity (Figure 22C). Vpr-RT-IN-PL is cotransfected with VSV-G and NLlucΔBglΔRI provirus into 293T cells to produce single cycle IN-PL virus (Figure 22C). C8166 T cells were first infected with the same amount of normal IN-PL virus or virus produced from shKu70- knockdown 293T cells (normalized by p24 ELISA) and

washed thoroughly at 3 h post-infection and collected half of the cells. The rest of cells were harvested at 8 h post-infection. All the cells were lysed and measured for IN-ProLabel activity. The result showed that when C8166 T cells were infected with virus produced from normal 293T cells, IN level remains unchanged after 8 h post-infection, as compared to that at 3 h post-infection (Figure 22C, lower panel, compare bar 2 to bar 1). When Ku70- knockdown C8166 T cells were infected with virus produced from normal 293T cells, IN level remained 77% after 8 h post-infection, as compared to that at 3 h post-infection (Figure 22C, lower panel, compare bar 4 to bar 3). Remarkably, When Ku70- knockdown C8166 T cells were infected with virus produced from Ku70-knockdown 293T cells, at 8 h post-infection, the IN level was reduced to approximately 34%, as compared to that at 3 h post-infection (Figure 22C, compare bar 6 to bar 5). All these results together suggest that Ku70 present in the progeny virus and in the target cells contribute to stabilizing IN in the early stage of HIV-1 replication.

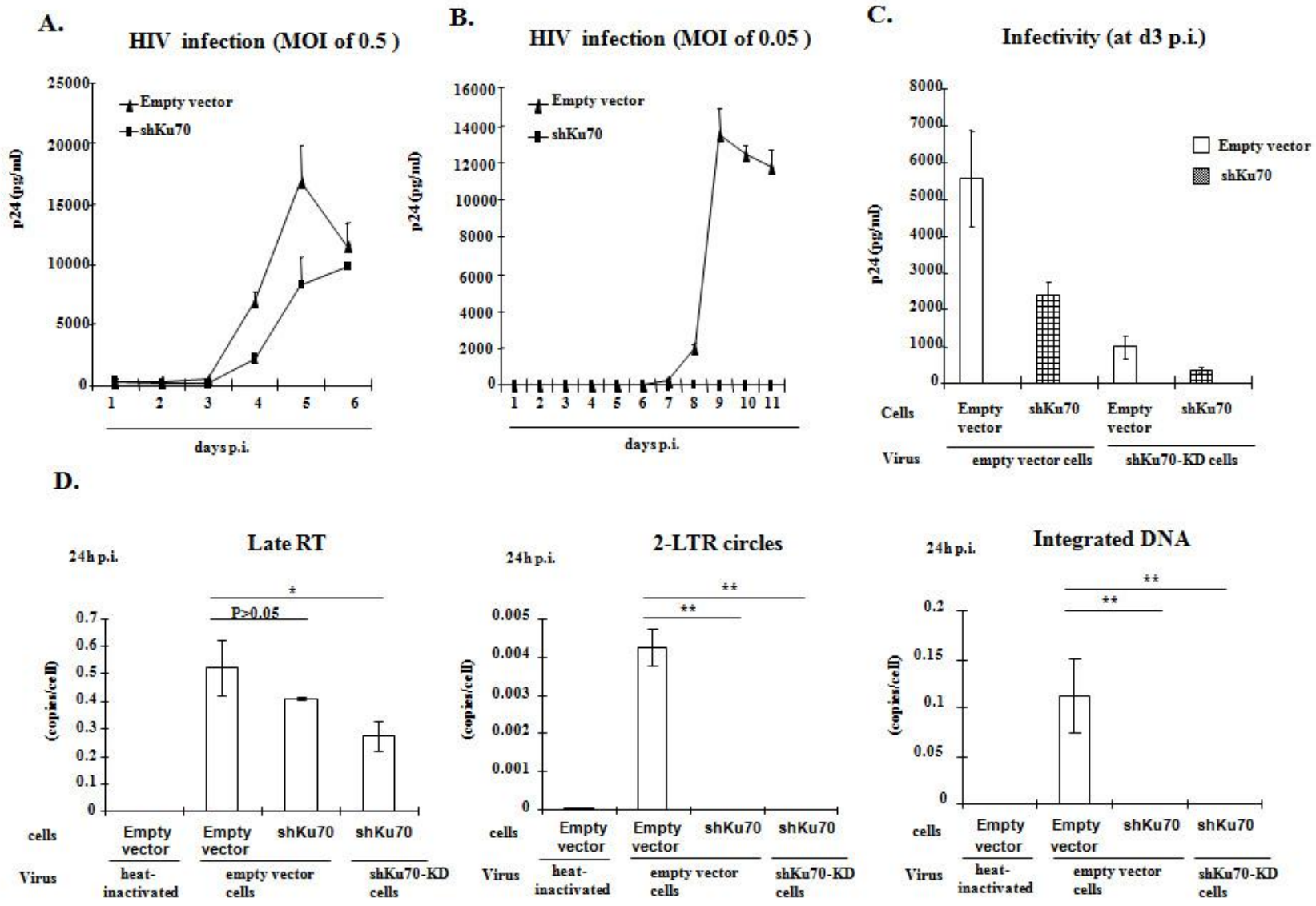
A.
HIV infection (MOI of 0.5)
p24 (pg/ml)
Empty vector
shKu70
days p.i.
B.
HIV infection (MOI of 0.05)
p24 (pg/ml)
Empty vector
shKu70
days p.i.
C.
Infectivity (at d3 p.i.)
p24 (pg/ml)
Empty vector
shKu70
Cells
Empty vector
shKu70
Empty vector
shKu70
Virus
empty vector cells
shKu70-KD cells
D.
Late RT
24h p.i.
(copies/cell)
P>0.05
*
cells
Empty vector
Empty vector
shKu70
shKu70
Virus
heat-inactivated
empty vector cells
shKu70-KD cells
2-LTR circles
24h p.i.
(copies/cell)
**
**
cells
Empty vector
Empty vector
shKu70
shKu70
Virus
heat-inactivated
empty vector cells
shKu70-KD cells
Integrated DNA
24h p.i.
(copies/cell)
**
**
cells
Empty vector
Empty vector
shKu70
shKu70
Virus
heat-inactivated
empty vector cells
shKu70-KD cells

Figure 21 Differential replication kinetics in Ku70-KD C8166 T cells with different titers of viral infection. A, B) HIV-1 replication kinetics in Ku70-KD or empty-vector-transduced C8166 T stable cell lines at a high MOI of 0.5 (A) or a low MOI of 0.05 (B). Lentiviral shRNA targeting Ku70 or empty-vector-transduced C8166 T stable cell lines were infected with different doses of pNL4.3-GFP virus for 2 h. At subsequent time intervals, the supernatants were collected, and viral replication was monitored by measuring HIV-1 p24gag levels. The data shown are the means and standard deviations of p24 values from duplicate wells in one infection assay and are representative of three independent experiments. C) Ku70 KD inhibited the infectivity of progeny virus. pNL4.3-GFP viruses produced from empty-vector-transduced or shKu70 KD C8166 T cells were normalized by p24 and used to infect both empty-vector-transduced and shKu70 KD C8166 T cells. Viral replication was monitored by p24 ELISA at three days post-infection. D) Ku70 KD impaired 2-LTR formation and integration of proviral DNA. Stable Ku70-KD or empty-vector C8166T cells were infected with the same amount of pNL4.3-GFP viruses produced from either empty-vector-transduced C8166T cells or shKu70-KD cells for 24 h. The cellular genomic DNA was extracted and quantified for HIV-1 late RT, 2-LTR and integrated DNA by real-time PCR. Infection with 70°C heat inactivated virus was served as a negative control (Bar 1). Data are representative of two independent experiments performed in duplicate, shown as mean ± SD. The statistical significance is denoted as * for p value ≤ 0.05 and ** for highly significance with p value ≤ 0.01. KD: knockdown.

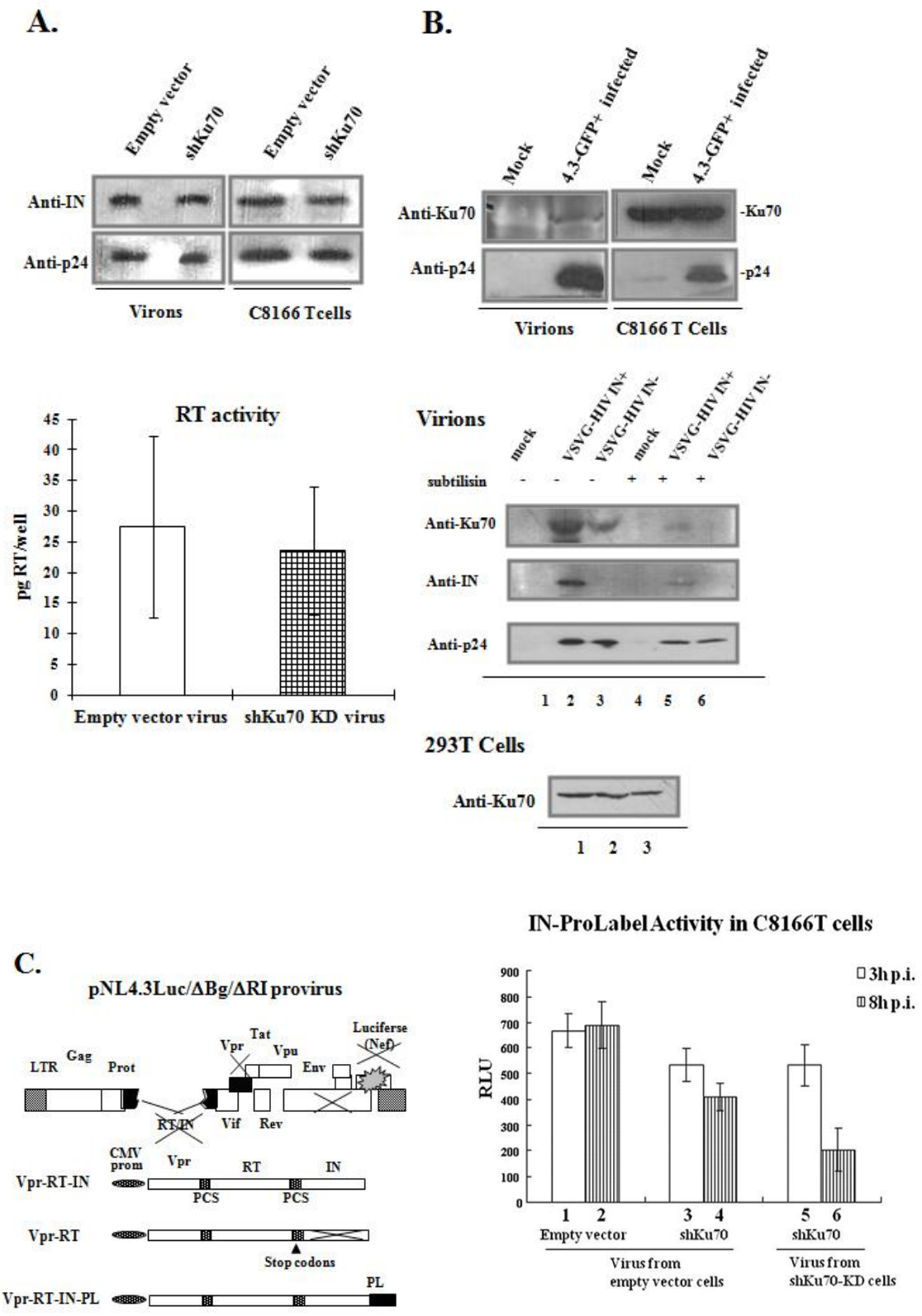

Figure 22 Ku70 incorporation into HIV-1 particles and its effects on HIV-1

replication. A) Knockdown of Ku70 does not affect the p24/IN profile in the virions or virus-producing cells. HIV pNL4.3-GFP virus produced from empty-vector-infected and shKu70-KD C8166 T cells were pelleted through a 20% sucrose cushion and dissolved in RPMI-1640. Viral particles with the same amounts of p24 and equal amounts of infected cells were lysed, separated by SDS-PAGE and immunoblotted with anti-IN and anti-p24 antibodies (upper panel). The data shown represent two independent experiments. The same amount of pNL4.3-GFP virus (normalized by p24) produced from empty-vector-transduced and shKu70 KD C8166 T cells were analyzed for HIV reverse transcription activity with a Reverse Transcriptase Assay kit (Roche). The data are shown as the amount of RT in each well, reported in pg/well, representing the means and standard deviations from viruses produced in two independent experiments (lower panel). B) Upper panel: Ku70 is present in the HIV-1 virion. C8166 T cells were infected with pNL4.3-GFP virus or uninfected. After four days of infection, supernatants were centrifuged through a 20% sucrose cushion at 35,000 rpm for 2 h at 4°C. Pellets were dissolved in RIPA buffer and subjected to a TCA-precipitation assay. Protein contents in the viruses and the cells were analyzed by WB using anti-Ku70 and anti-p24 antibodies to determine the presence of various proteins. Lower panel: Ku70 incorporation into HIV-1 virion is dependent on IN. The Vpr-RT-IN or Vpr-RT expressor was cotransfected with VSV-G and NL4.3lucΔBglΔRI to produce single cycle IN+ and IN- virus. The viruses were subjected to Subtilisin resistance assay as described in Material and Methods. Virus-associated Ku70, IN and p24 were analyzed by WB. Endogenous Ku70 expressions in the transfected 293T cells were checked by blotting with anti-Ku70 antibody. C) Upper panel: schematic structure of RT/IN/Env deleted HIV-1 proviruses

NL4.3lucΔBglΔRI and of the Vpr-RT-IN, Vpr-RT, Vpr-RT-IN-PL fusion protein. NL4.3lucΔBglΔRI and Vpr-RT-IN were described earlier [282], and Vpr-RT without IN expression has two stop codons TAGTGA after last nucleotide of RT sequence. Vpr-RT-IN-PL was obtained through removal of IN stop codon and insertion of ProLabel gene after IN of Vpr-RT-IN plasmid as described in Material and Methods. Lower panel: NL4.3lucΔBglΔRI was cotransfected with Vpr-RT-IN-PL and VSV-G expressor into 293T cells to generate VSV-G-pseudotyped HIV-1 single cycle IN-PL virus. 2×10^6 shKu70-KD or empty-vector-transduced C8166 T cells were infected with of IN-PL virus (2000 pg p24) produced from shKu70-KD or normal empty-vector-transduced 293T cells for 3 h. The cells were washed three times and half of the cells were collected. The rest of cells were kept in RPMI medium and harvested at 8 h post infection. All the cells were then lysed and measured for ProLabel activity by the POLARstar OPTIMA multidetection microplate reader. The results are representative of three experiments, shown as mean ± SD. RLU: relative light unit; KD: knockdown.

6.5 Discussion

We studied the interaction between the cellular DNA-repair protein Ku70 and HIV-1 IN, and the potential roles of Ku70 in HIV-1 replication. By using a cell-based Co-IP assay, we demonstrated that Ku70 interacts with HIV-1 IN in 293T cells and HIV-1 infected CD4+ T cells. Deletion analyses on IN and Ku70 indicated that an IN region encompassing aa 1–230 was unable to bind to Ku70, whereas the N-terminal region of Ku70 (aa 1–430) still retained the IN binding ability, and that their interaction is independent of Ku70/80 heterodimerization. We further discovered a dual mechanism for Ku70 in protecting IN from proteasomal degradation, i.e., by reducing overall protein ubiquitination levels within the host cells and by specifically reducing the ubiquitination of IN via their binding interaction. Finally, the knockdown of Ku70 expression in both virus-targeting and virus-producing CD4+ T cells significantly impaired HIV-1 replication. More specifically, Ku70 knockdown resulted in undetectable 2-LTR and integration levels in the early stage of viral replication. Taken together, our current study suggests that Ku70 is required for both the early and late stages of the HIV life cycle.

6.5.1 Mechanism of protection by Ku70 against host UPS in IN

A recent study has found that Ku70 is able to reduce the ubiquitination of Bax in the regulation of apoptosis [323]. This study indicated that the presence of Ku70 is able to specifically deubiquitinate Bax; however, whether Ku70 could affect total ubiquitination level in the host cells remained unknown. Presently, we discovered that, besides reducing the ubiquitination of Bax, Ku70 is able to universally down-regulate the ubiquitination of the entire complement of cellular proteins. The mechanism underlying

this down-regulation of protein ubiquitination levels by Ku70 is unclear. A prior *in vitro* study revealed that Ku70, defined as a novel DUB enzyme, was able to hydrolyze polyubiquitin chains into monoubiquitin units [323]. In the cascade of the ubiquitin-proteasome pathway, DUBs remove ubiquitin chains from protein substrates and recycle the polyubiquitin chain into free ubiquitin, before or after the polyubiquitination chain is recognized by the 19S cap of the 20S proteasome [331]. Thus, if Ku70 does exert an overall DUB activity on cellular protein, an increased free ubiquitin level in the cells would be expected. Unfortunately, under our experimental conditions, we were not able to detect an increased monoubiquitin form of ubiquitin when Ku70 was overexpressed. It appears that the total pool of monoubiquitin in the cells was reduced when Ku70 was overexpressed. Thus, how Ku70 interacts with the host UPS to regulate the total pool of monoubiquitin in the cells and down-regulate the polyubiquitination of proteins remains an open question.

During the HIV life cycle, the host UPS is repeatedly used by HIV-1 viral proteins such as Vif, Vpr, Vpu and IN to ensure viral replication, either by targeting host restriction factors for degradation, such as Vif/APOBEC3G or Vpu/BST2, or by protection from proteasomal degradation by host cellular cofactors (e.g., IN is protected by LEDGF/p75 and Vpr is protected by Cul4A-DDB1DCAF1 ubiquitin ligase) [42, 43, 175, 332-334]. Here, we provide another example of a host cellular cofactor of IN, the DNA-repair protein Ku70, which protects IN from host proteasomal degradation in overexpression system and under HIV-1 infection (Figure 17, 20B upper panel and Figure 22C lower figure). Interestingly, our results showed that, in addition to down-regulating the ubiquitin pool within the cells, Ku70 is also able to specifically reduce the ubiquitination

level of IN and its associated proteins through IN/Ku701-430 binding. Indeed, the IN-binding peptide Ku701-430 was able to increase IN expression and reduce the ubiquitination of IN (and its associated proteins) even though it does not have any DUB activity (Figure 20B; compare lane 4 to lane 2 of the first and second panels). This suggests that Ku70 specifically binds IN and possibly masks ubiquitin-attachment site(s) in IN to reduce the ubiquitination of IN, thus protecting IN from degradation (Figure 20B, upper panel). This scenario seems very likely, as the attachment of ubiquitin to protein substrates, conducted by ubiquitin ligases E3 during the last step, is via the covalent binding of the C-terminal G76 of ubiquitin to the ε-amino group of an internal lysine residue in the substrate protein [335], and the C-terminus of IN (230–288) is known to be enriched in lysine residues (e.g., K240, K244, and K264). Thus, in a future study, it would be interesting to investigate which lysine residue(s) in IN are specifically recognized by ubiquitin and subjected to proteasomal degradation.

6.5.2 Ku70: a newly identified cellular cofactor of HIV-1 IN

We also provided solid evidence that HIV-1 IN directly interacts with Ku70 in Ku70/IN-overexpressing mammalian cells and in T-lymphocytes during HIV-1 infection (Figure 18). Indeed, this finding is not surprising considering that previous studies using M-MLV IN as bait in a yeast two-hybrid system were able to fish out Ku70 [150], that Ku70 is present in M-MLV PICs [211] and that Ku70 associates with the IN of Ty elements in *S. cerevisiae*, yeast retrotransposons with life cycles similar to retroviruses [336]. We then extended our study to delineate the mutual binding interface of IN and Ku70 and found that the C-terminus of IN (aa 230–288) appears to be involved in the interaction with Ku70. One interesting observation here is that the Ku70 deletion mutant

1–430, which was able to bind IN, cannot form the Ku70/80 heterodimer (Figure 19). This result is consistent with a previous finding from a two-hybrid analysis that the C-terminal 20 kD of Ku70 (aa 430–609) is essential for Ku70/80 heterodimerization [182], and suggests that the Ku70/IN interaction may be independent of Ku80. It has been shown that Ku70 has unique functions independent of Ku80, although many cellular functions in which Ku70 participates do require Ku70/80 heterodimerization. For example, the antiapoptotic activity of Ku70 by inhibiting Bax-mediated apoptosis is independent of Ku80 [337]. One concern here is that, despite the fact that the Ku70 N-terminus (aa 1–430) can still mediate IN binding, we cannot rule out the possibility that IN is able to bind the Ku70/80 heterodimer or that IN binding to Ku70 may be enhanced by Ku70/80 association in mammalian cells. Indeed, several cellular functions of Ku70 require association with Ku80, and this heterodimerization formation enhances the stability of each subunit [338, 339].

6.5.3 Requirement of Ku70 in both of the early and late stages of HIV-1 life cycle

In a finding of major relevance, we further determined that Ku70 is required for HIV-1 replication. HIV-1 infection was significantly impaired when Ku70 expression was knocked down in both producer and target cells (Figure 21). Moreover, with a low-MOI infection, Ku70- knockdown C8166 T cells were more significantly blocked in viral replication than for a high-MOI infection (Figure 21 A and B). These results indicate that Ku70 affects both the early and the late steps of the HIV-1 life cycle. In order to pinpoint the effect(s) of Ku70 on the early stage, we carried out real-time PCR to quantify late RT, 2-LTR circles and integrated DNA under the same infection condition as Figure 21C. Not surprisingly, 2-LTR circles formation in the Ku70- knockdown cells

was undetectable which is consistent with previous report that NHEJ pathway is required for 2-LTR circle formation (Figure 21D) [211, 329]. In addition, integration of viral DNA was also abrogated by Ku70 knockdown, while reverse transcription measured by Late RT products reduced by around 50% in the Ku70- knockdown cells infected with sh-Ku70 virus produced from Ku70- knockdown cells (Figure 21D). Thus, Ku70 seems to be a multi-faceted player in the early stage of viral infection. Ku70 is a component of the NHEJ pathway [195] that has been extensively investigated for its multiple functions during retroviral transduction or infection in previous studies. With respect to the effects of Ku70 on the early stage of HIV replication, one possible mechanism might be through the NHEJ pathway, such as the protection of infected cells from apoptosis, circularization of the viral DNA, or the gap repair of integration intermediates introduced by HIV-1 IN [204, 206-209, 211, 340]. These scenarios explain undetectable 2-LTR circyles and integration event when Ku70 is absent during viral infection. However, whether the NHEJ pathway is strictly required for 2-LTR and integration was challenged by the study done by Kilzer *et al* who reported that another NHEJ pathway component DNA-PKcs had limited effect on 2-LTR formation and integration [329]. While the controversy could result from different experimental methods used to quantify 2-LTR and integration and different cell lines, more detailed studies are surely needed to fully understand the roles of Ku70 on DNA circularization and integration. The second possible mechanism, with a plausibility demonstrated in this study, is that Ku70 may protect IN from degradation before viral DNA integration by reducing overall cellular ubiquitination and/or by specifically interacting with IN in HIV-infected cells. This scenario was supported by studying the effect of Ku70 on IN-PL metabolism after viral entry in the

early stage of viral replication (Figure 22C). Depletion of Ku70 in the target cells had reduced IN expression to 77% at 8 h post-infection. Remarkably, when Ku70-knockdown C8166 T cells were infected with virus produced from Ku70- knockdown 293T cells, the IN level at 8 h post-infection, was reduced to approximately 34%, as compared to that at 3 h post-infection (Figure 22C, lower panel, compare bar 6 to bar 5). These observations imply that both Ku70 associated with virus and the cells are able to protect IN to avoid host ubiquitin proteasomal degradation in the PIC. At this point, it should be noted that two other cellular proteins, hRad18 and LEDGF/p75, were also previously reported to protect IN from proteasomal degradation [175, 226]. However, unlike these two IN cofactors, whose actions are mainly through their physiological binding to IN, Ku70 instead displays two different activities to protect IN from degradation. To date, it has not been clear how HIV-1 IN can coordinately recruit these cofactors to protect itself from host proteasomal degradation machinery.

With regard to the effect of Ku70 on the late stage of the viral life cycle, we originally hypothesized that Ku70 knockdown might affect the Gag/Gag-Pol ratio, as Ku70 might protect IN as an early Gag-Pol polyprotein precursor during viral assembly. In fact, the results shown in Figure 22A demonstrate that Ku70 does not affect Gag/Gag-Pol processing and maturation. Further analysis of virion composition revealed that the host cellular protein Ku70 is present in the HIV-1 particles themselves (Figure 22B). This finding indicates that Ku70 is packaged into HIV-1 particles as early as its assembly stage and becomes part of the HIV-1 PICs after the virus enters target cells. Within the PIC and associated with IN, Ku70 might deploy two mechanisms to contribute to the early stage of HIV replication: 1) protecting IN from the host proteasomal degradation

pathway and 2) assisting viral protein IN in specific replication step(s) including 2-LTR formation and integration. Consistently, our data revealed that the progeny virus produced from Ku70- knockdown cells was profoundly defective even when they were used to infect normal cells (Figure 21C, bar 3). However, we also cannot exclude the possibility that the presence of Ku70 is required for other events during HIV-1 morphogenesis. This notion is strengthened by the previous findings that, during infection, free ubiquitin is incorporated into viral particles of HIV-1, simian immunodeficiency virus (SIV), M-MLV and equine infectious anemia virus (EIAV) [341] and that pr55Gag is monoubiquitinated during assembly and budding [342]. In addition, proteasome-inhibitor treatment interfered with the assembly and budding of HIV progeny virus and efficiently inhibited HIV infectivity [343]. Thus, further investigation is certainly needed to fully understand how Ku70 impacts HIV-1 replication, and a better understanding of the interplay between HIV-1 IN and Ku70 during viral infection will rationalize the design of specific inhibitors to target this viral-cellular protein interaction and consequently inhibit HIV-1 replication.

Chapter 7

Roles of putative SUMO-interaction motifs (SIMs) of HIV-1 integrase in mediating its cofactor binding and nuclear import

7.1 Rational

Small Ubiquitin-like Modifier or SUMO proteins are covalently attached to protein substrates through the canonical four-amino-acid SUMOylation consensus sites ψ-K-x-D/E (or SUMO-CS, ψ is an aliphatic branched amino acid and x is any amino acid). SUMOylation affects various cellular processes, such as protein stability, localization, DNA binding, and activation (See a review [234]). A recent study has revealed that HIV-1 IN contains three SUMOylation sites (45LKGE, 135IKQE and 243WKQE) at three Lys residues (K46, K136 and K244) and SUMOylation of IN functions in the early stage of infection, before integration but after reverse transcription [214]. When these three critical Lys residues K46, K136 and K244 were mutated into similar positive charge Arg, SUMO modification of IN was drastically reduced, suggesting that these three amino acids are major SUMOylation sites [214]. However, SUMOylation of IN is not completely abolished with all the mutations, highlighting the presence of other SUMOylation sites. This suggested that SUMOylation of IN might take place on the Lys residues within non-consensus SUMOylation region or inverted SUMOylation consensus motif E/DxKψ, as reported in other proteins such PCNA and human E2-25K [232, 243, 244]. Furthermore, SUMOs have three major isoforms (SUMO1-3), all of which have been shown to modify IN in the *in vitro* and *in vivo* studies [214]. However, it is still unclear with respect to which SUMO subtype(s) preferentially targets IN.

In addition to covalent SUMO conjugation, SUMO proteins also non-covalently bind proteins through SIM. The best characterized of SIM sequence is V/I-x-V/I-V/I or V/I-V/I-x-V/I/L (where x can be any amino acid) [4, 5]. Of note, SUMO proteins from SUMO-modified proteins provide platforms for non-covalent binding to SIM-containing binding partners or itself (Figure 5C and 5D). The functional consequences of SIM-SUMO interaction vary considerably in different protein contexts, affecting SUMOylation, recruitment, localization, protein conformation, etc. (see a review [242]). Interestingly, a few IN-interacting proteins such as LEDGF/p75, Ku70, p300 and Rad52 are also SUMOylated [246-249]. Thus, it may be easily conceived that the SIM-SUMO interaction might mediate the complex formation between SIMs of IN and its SUMOylated cofactors and affect different functions of IN. Based on this consensus SIM, we identified three putative SIMs (72VILV75, 200IVDI203 and 257IKII260) in HIV IN in this study. The focus of this section was to answer the questions of whether IN bears SIMs and whether IN recruits its SUMOylated cofactors through SIM and its functions.

7.2 Hypothesis

In this study, we hypothesize that these IN putative SIMs play various roles including SUMOylation, recruitment of cofactors, nuclear import of IN and impact on HIV-1 infection.

7.3 Objectives

1. Investigate the roles of these IN putative SIMs in SUMOylation of IN
2. Study the regulation of putative SIMs of IN on the binding of IN-LEDGF/p75 and IN-Ku70

3. Evaluate the effect of IN putative SIMs on its nuclear translocation.
4. Assess the SUMO binding ability of IN and its impacts on HIV-1 replication

7.4 Results

7.4.1 HIV-1 IN contains three putative SIMs

SUMOylation of IN was identified in 2011 and three SUMO conjugation sites (45LKGE, 135IKQE and 243WKQE) (Figure 23A) were shown to be SUMOylated in the *in vivo* and *in vitro* SUMOylation assay [214]. However, the functional importance of this PTM in the roles of IN during viral replication remains largely unknown. An examination of the amino acid sequence of IN indicated that it harbours three putative SUMO binding sites or SIMs. The four-aa-hydrophobic cores of these SIMs in IN are: SIM1 72VILV75; SIM2 200IVDI203 and SIM3 257IKII260. Sequence alignment of these three SIMs of IN with SUMO E2 protein RANBP2 and PIAS1 revealed that these three SIMs fit into the consensus with the four-aa-hydrophobic core V/I-x-V/I-V/I or V/I-V/I-x-V/I/L (Figure 23B). To determine the importance of these sites to IN's functions, we independently mutated three SIMs: the first SIM (M1 mutant), the second SIM (M2 mutant), and the third SIM (M3 mutant). And we also mutated triple SIMs mutant 3VI and the combination of any two SIMs mutants denoted by M1+M2, M1+M3 and M2+M3 (Figure 23C, not shown for M1+M2, M1+M3 and M2+M3). Meanwhile, we mutated three Lys in the consensus SUMO conjugation sites (K46/K136/K244) into Arg and this SUMOylation-defective mutant refers to 3KR.

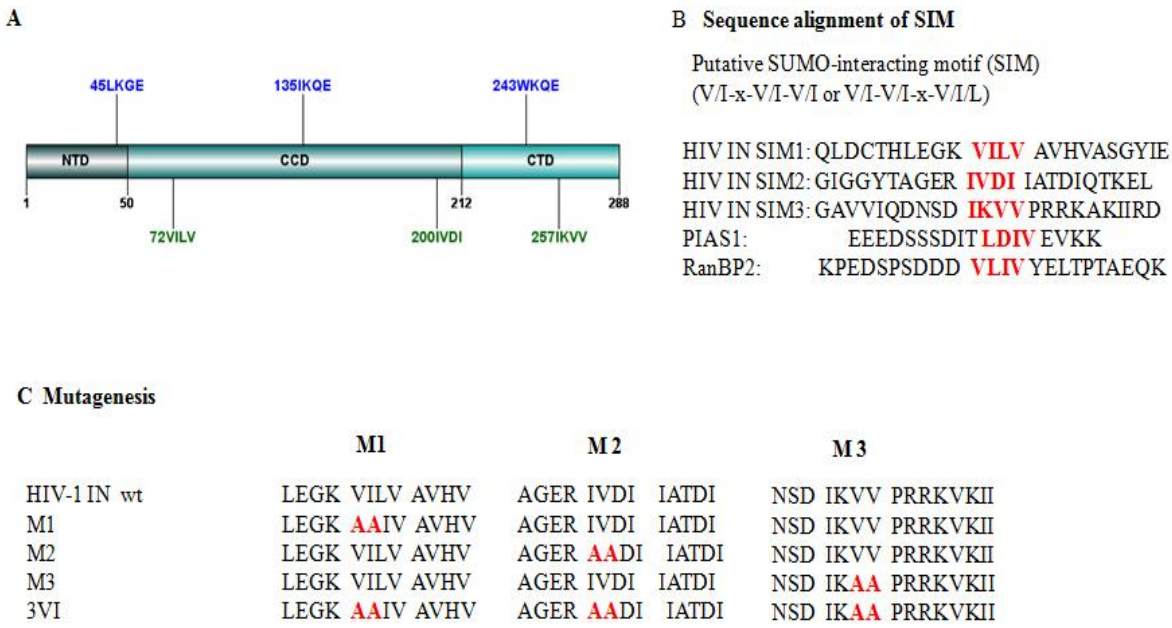

Figure 23 IN harbors three putative SIMs. A) Schematic of HIV-1 IN with its three putative SIMs and SUMO conjugation sites highlighted. IN contains three ψ-K-x-E SUMO conjugation sites (K46, K136 and K244) [214] and three putative SIMs. The core of the SIM is composed of three hydrophobic (I, L or V) residues, arranged as V/I-V/I-x-V/I/L or V/I-x-V/I-V/I (where x can be any amino acid). Four residues that constitute the hydrophobic core of each SIM are indicated (SIM1 72VILV75; SIM2 200IVDI203 and 257IKII260). B) Sequence alignment of three IN SIMs with SUMO E3 protein RANBP2 and PIAS1, with the core highlighted in red. C) Mutational analysis of IN putative SIMs. The amino acids of IN SIMs mutated are shown in red. Ala had been introduced into the SIM1, SIM2, SIM3 or three SIMs to generate SIM1 mutant M1, SIM 2 mutant M2, SIM3 mutant M3, or three of SIMs mutant 3VI. SIM: SUMO-interacting motif.

7.4.2 Three putative SIMs present in IN disfavour its own SUMOylation

SUMO binding has been shown to regulate protein SUMOylation (see review [242]). To study the effect of SUMO binding of IN in its SUMOylation, SUMOylation of IN was first confirmed by *in vivo* SUMOylation assay as described in Material and Method (Figure 24A). To do so, HA-SUMO3 and GFP-INopt wt was coexpressed into 293T cells. The SUMO conjugated IN was detected by immunoprecipitating with anti-GFP antibody and immunoblotting with anti-HA antibody (Figure 24A, upper panel). Meanwhile, expression levels of GFP-INopt wt and HA-SUMO3 in the cells were verified by WB (Figure 24A middle and lower panel). To further verify whether SUMO binding ability of IN regulates its SUMOylation, SUMOylation of IN wt, 3VI and 3KR was detected by using the same *in vivo* SUMOylation assay. The IN mutant 3KR which showed reduced SUMOylation in the previous report [214] is also included in this study. Unexpectedly, the result showed that the triple SIMs mutant 3VI increased SUMOylation of IN by four-fold while previously identified SUMOylation-defective mutant 3KR had the same SUMOylation level with IN wt (Figure 24B). We had repeatedly detected increased SUMOylation of 3VI and intact SUMO3 conjugation of IN mutant 3KR under our experimental condition.

Ubc9 is the unique SUMO E2 conjugating enzyme required for all SUMOs conjugation, or SUMOylation. Many of the proteins modified by SUMO directly interact with Ubc9, such as IκBα, p53, and RanGAP1, PCNA, Rad18, Rad5 [244, 344, 345]. And this interaction between Ubc9 and target proteins is important for substrate recognition and substrate conjugation. Studies have shown that mutations within the ψKxE conjugation sites abolished both Ubc9 binding and substrate modification [344]. To investigate the

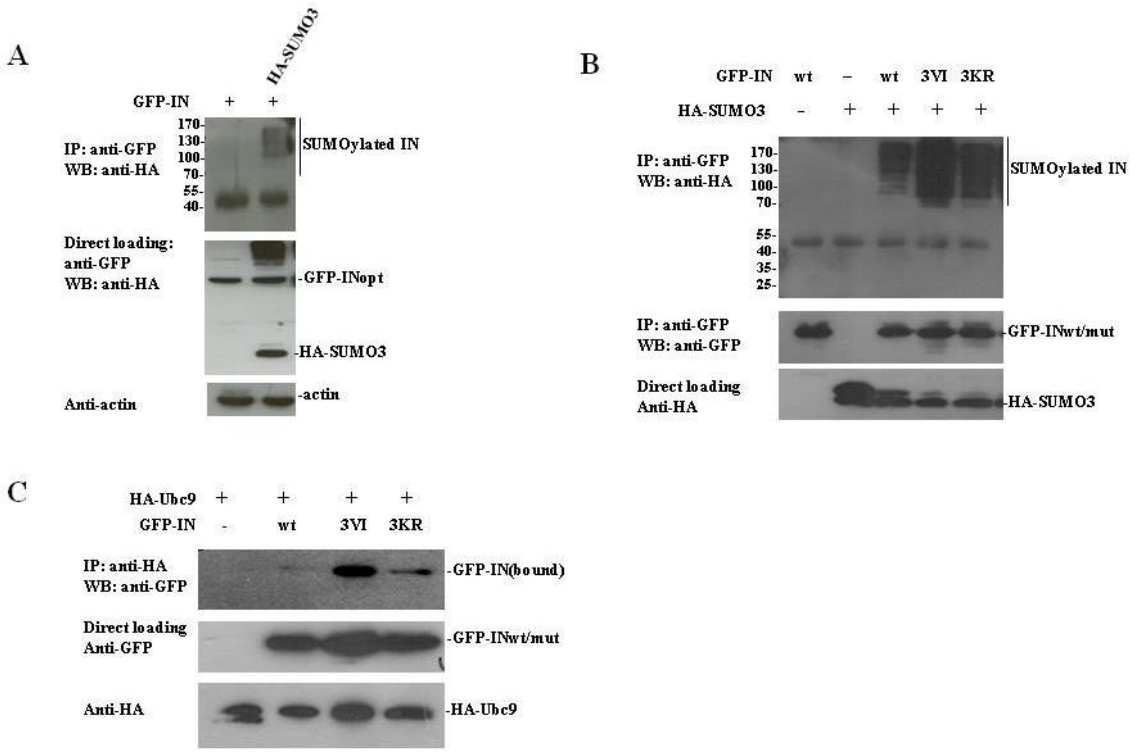

Figure 24 Putative SUMO binding of IN disfavor its SUMOylation. A) IN is SUMOylated in the *in vivo* SUMOylation assay. 293T cells were cotransfected with GFP-INopt and HA-SUMO3 or GFP-INopt alone. SUMOylation of IN by HA-SUMO3 is detected by *in vivo* SUMOylation assay as described in Material and Methods. At 40 h post-transfection, the lysates were immunoprecipitated with anti-GFP antibody, followed by WB analysis using antibody against HA to detect SUMOylation level of IN (upper panel). Meanwhile, 5% of total cells were lysed in 4 x Laemmli Buffer and loaded onto SDS-PAGE gel to detect GFP-INopt and HA-SUMO3 expression (middle panel). In the lower panel, endogenous beta-actin was detected as a loading control. B) 293T cells were co-transfected with different plasmids as indicated and SUMOylation of GFP-INopt wt and mutants were analyzed by using the same *in vivo* SUMOylation assay. SUMOylation level of INwt/mut is shown in the top panel. Expression levels of GFP-INwt/mut and HA-SUMO3 are shown in the middle and lower panel, respectively. C) The interaction

between IN and SUMO E2 protein Ubc9 was increased in IN SIMs mutant 3VI. HA-Ubc9 was coexpressed with various GFP-INopt wt/mut plasmids into 293T cells and subjected to Co-IP assay to study the binding affinity of Ubc9-IN. The cell lysates were immunoprecipated with anti-HA antibody to pull down HA-Ubc9-bound GFP-IN using anti-GFP antibody in the WB (upper panel). Meanwhile, 5% of total cells were lysed to check expressions of GFP-INopt wt/mut and HA-Ubc9 (middle and lower panel).

potential roles of Ubc9 in the SUMO binding of IN, we studied the interaction between HA-Ubc9 and GFP-INopt wt, 3VI and 3KR mutants. Co-IP data revealed that HIV-1 IN wt and 3KR had the similar level of binding affinity with HA-Ubc9, while 3VI mutant had greatly increased binding with HA-Ubc9 compared with wt and 3KR (Figure 24C, upper panel). The increased 3VI/HA-Ubc9 binding is consistent with increased SUMOylation level of 3VI (Figure 24B). The observation that IN mutant 3VI had enhanced SUMOylation and Ubc9 binding suggested that SUMO binding of IN might disfavor its own SUMOylation. Given that the SUMOylation consensus motif ψKxE is a major determinant of Ubc9 binding and SUMO modification, and IN 3KR mutant shared similar levels of SUMO conjugation and HA-Ubc9 binding with IN wt (Figure 24B. and Figure 24C), our data reinforced the previous observation that 3K (K46/K136/K244) of IN are not the major conjugation sites for SUMO3 modification *in vivo*.

7.4.3 Roles of putative SIMs of IN in its cofactor(s) binding

SUMO-binding or SIM sequence is able to mediate protein-protein interactions through the binding between SUMOylated protein and SIM-containing binding partner (Figure 5C). For example, the SIM of RanBP2/Nup358 is responsible for the interaction between RanBP2/Nup358 and SUMOylated RanGAP1 [240]. To investigate whether putative SIMs of IN could mediate the interaction with its SUMOylated cofactors, we have studied the binding ability of IN-LEDGF/p75 and IN-Ku70 as both LEDGF/p75 and Ku70 are SUMOylated in previous studies [246, 346]. First, we investigated the interaction of IN-LEDGF/p75 by Co-IP assay, in which 293T cells were co-transfected with T7-LEDGF/p75 and GFP-INopt wt or IN mutants 3VI, 3KR. Results showed that IN 3KR mutant still bound T7-LEDGF/p75 to the same extent with IN wt which is

consistent with the previous report (Figure 25A) [214], while 3VI mutant completely lost binding ability with T7-LEDGF/p75 (Figure 25A, upper panel). To investigate which SIM(s) within IN might be essential for LEDGF/p75 binding, we individually or simultaneously mutated these three putative SIMs and the mutants including 3VI, M1, M2, M3 and M1+M2, M1+M3, M2+M3 were co-tranfected with T7-LEDGF/p75 into 293T cells. The results obtained from Co-IP assay showed that IN M1 mutant still retained full LEDGF/p75 binding ability as IN wt, whereas LEDGF/p75 binding affinities were compromised to different extents in other IN SIM mutants (3VI, M2, M3, M1+M2, M1+M3, M2+M3) (Figure 25B).

Since mutations within the putative SIMs of IN impaired LEDGF/p75 binding, it was hypothesized that these motifs of IN might mediate interaction with SUMO conjugated LEDGF/p75. Four SUMOylation sites (K75, K250, K254 and K364) of LEDGF/p75 were revealed in the previous study [346]. To test this hypothesis, these four Lys were first mutated into Arg to abolish its SUMOylation ability; the resulting mutant was denoted LEDGF/p75 4K mutant [346]. Then, T7-LEDGF/p75 wt and 4K mutant was individually cotransfected with GFP-INopt wt into 293T cells to study their interactions by a cell-based Co-IP assay. Unexpectedly, the result showed that T7-LEDGF/p75 wt and 4K had similar binding affinity with GFP-INopt wt (Figure 25C, upper panel). Meanwhile, the similar expression levels of LEDGF/p75 and IN in each sample were detected as loading control (Figure 25C, middle and lower panel). Thus, the result suggests that these putative SIMs of IN do not directly impact on the binding of SUMO conjugated LEDGF/p75.

To study the potential effect of SUMO binding of IN in IN-Ku70 interaction, GFP-INopt

wt, 3KR, 3VI was individually co-transfected with Prolabel-Ku70 (PL-Ku70) or T7-Ku70 in 293T cells. The interaction of IN-Ku70 was verified by both of chemiluminescent Co-IP assay (Figure 26A) and Co-IP assay (Figure 26B). The results from two different detecting systems (chemiluminescent signal and autoradiography) were consistent. Results proved that while the interaction between IN mutant 3KR and T7-Ku70 or PL-Ku70 had similar binding ability with wt, 3VI increased binding affinity with Ku70 by two-fold (Figure 26A left panel, 26B upper panel). Meanwhile, co-transfection between GFP-C with PL-Ku70 or T7-Ku70 was included as a negative control. The expressions of various proteins including GFP, GFP-INopt wt/3KR/3VI, PL-Ku70, T7-Ku70 were detected by WB using corresponding antibodies (Figure 26A right panel, 26B middle and lower panel).

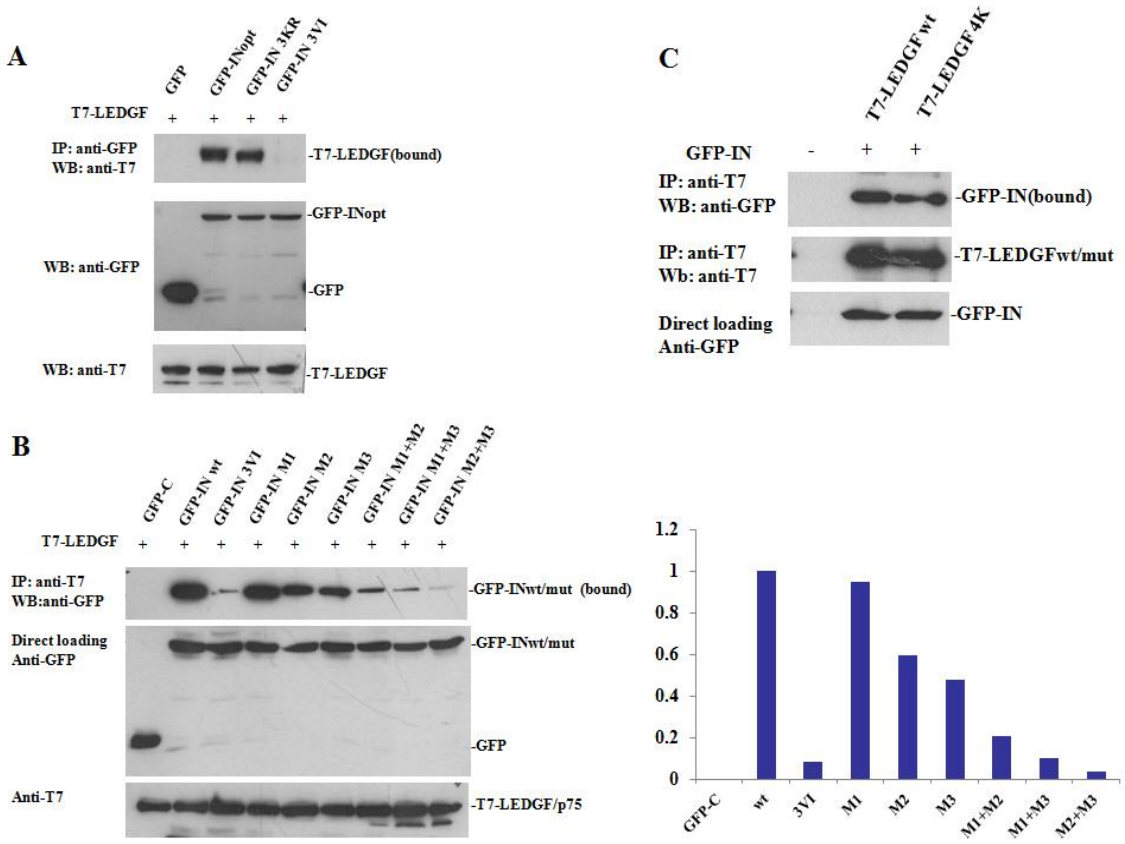

Figure 25 Putative SIMs of IN are indispensable for LEDGF/p75 binding. A) The interaction of T7-LEDGF/p75 with GFP-INopt wt or two mutants 3KR, 3VI was analyzed by Co-IP assay. The GFP-IN-bound T7-LEDGF/p75 is shown in the upper panel by immunoprecipating the cell lysates with anti-GFP antibody and immunoblotting with anti-T7-HRP antibody in the WB analysis. GFP alone was coexpressed with T7-LEDGF as a negative control. Expression levels of GFP, GFP-INopt wt/mut and T7-LEDGF/p75 are shown in the middle and lower panels. B) T7-LEDGF/p75 was cotransfected with single SIM mutant (M1, M2 and M3) or the double SIM mutants (M1+M2, M1+M3 and M2+M3) or triple SIM mutant 3VI into 293T cells for 40 h. The cell lysates were immunoprecipitated with anti-T7 antibody. Bound proteins were eluted with SDS-sample

buffer and analyzed by SDS-PAGE and autoradiography, detected by anti-GFP-HRP antibody (upper panel of the left Figure). GFP, GFP-INwt/mut and T7-LEDGF/p75 expressions are shown in the middle and lower panel of the left figure, respectively. The intensity of protein bands was densitometrically determined. Results were expressed as the ratio of bound GFP-INopt w/mut expression which was normalized by total input in the right figure. Binding affinity of GFP-INopt wt and T7-LEDGF/p75 was arbitrarily set as 100%. Results are representative of two independent experiments. C) SUMOylation defective LEDGF/p75 mutant 4K still binds IN. T7-LEDGF/p75 wt or 4K mutant was co-transfected with GFP-INopt wt in 293T cells and the interaction between GFP-IN and T7-LEDGF/p75 was analyzed by Co-IP assay at 40 h post-transfection. LEDGF/p75-bound GFP-INopt wt was detected by immunopreciptating with anti-T7 antibody and immunoblotting with anti-GFP-HRP antibody in WB (upper panel). The expression levels of transfected T7-LEDGF/p75 wt or 4K mutant and GFP-INopt wt were examined from the same amount of cell lysates (middle and lower panel).

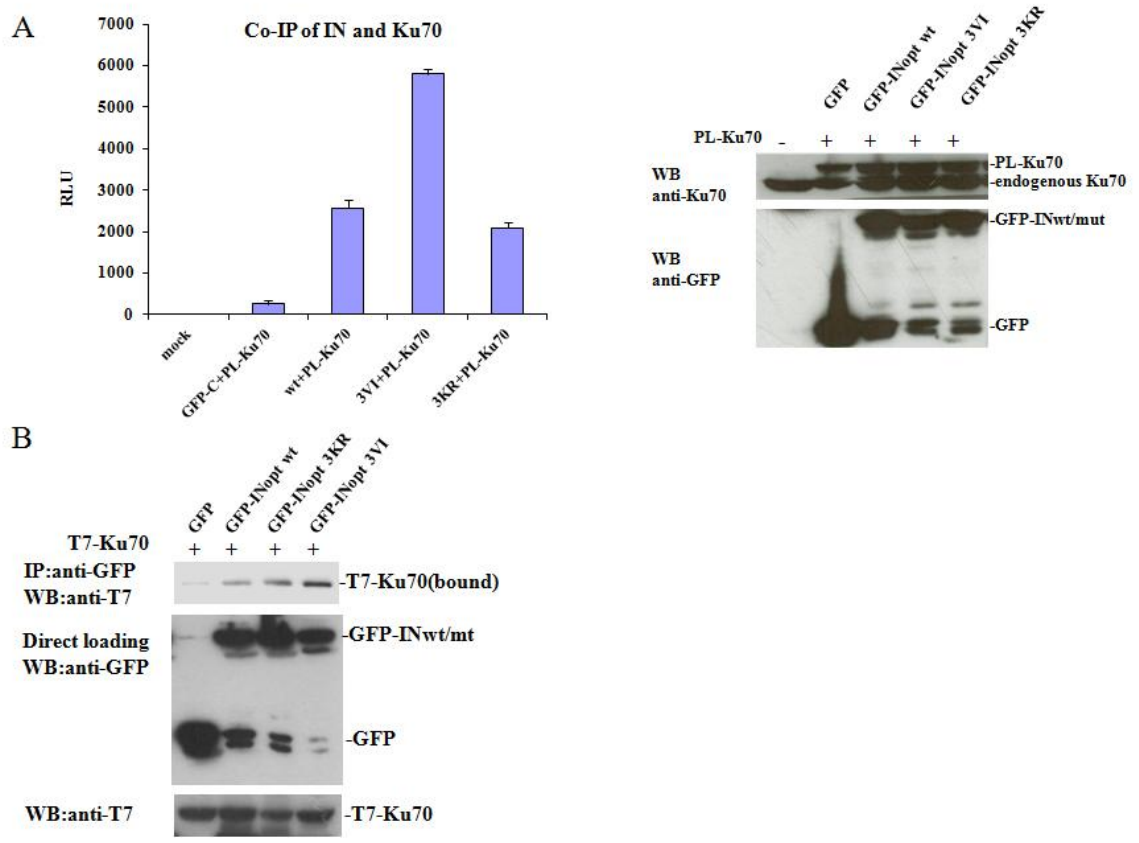

Figure 26 IN SIMs mutant 3VI had increased binding affinity with Ku70. A) The interaction of ProLabel-Ku70 (PL-Ku70) with GFP-INopt wt/mut was detected by chemiluminescent Co-IP assay. GFP-INopt wt/mut or GFP1-C and PL-Ku70 fusion proteins were coexpressed in 293T cells. After 40 h of transfection, cells were lysed and immunoprecipitated with anti-GFP antibody and the chemiluminescent signals from IN-bound PL-Ku70 present in the complexes were measured using ProLabel Detection Kit II and valued as relative luminescence units (RLU). Expression levels of PL-Ku70 and GFP-INopt wt/mut or GFP alone in each sample were analyzed by anti-Ku70 and anti-GFP-HRP antibodies (right panel). Results are representative of two independent experiments. B) 293T cells were co-transfected with SVCMVin-T7-Ku70 expressor and GFP-INwt/mut plasmid as indicated. After 40 h of transfection, 90% of cells were lysed

and subjected to Co-IP assay as described in Material and Methods. The upper panel showed the bound T7-Ku70 in each sample. 5% of cell lysates were used to detect the expression of GFP, GFP-INwt/mut and T7-Ku70 by WB using anti-GFP and anti-T7-HRP antibodies respectively (middle panel and lower panel).

7.4.4 Putative SIMs of IN is required for its nuclear localization.

Previous studies have revealed that SUMO-SIM interaction affects localization of SIM-containing proteins [347, 348]. To study whether these three putative SIMs of IN would play a role in its subcellular localization, HeLa cells were transfected with different IN plasmids GFP-INopt wt, 3KR, 3VI, M1, M2 and M3. The cells were then fixed, permeabilized and stained with anti-GFP antibody and FITC-labeled secondary antibody (green). The cellular nuclei were stained with DAPI (blue). The localization of IN wt and mutants were visualized by immunofluorescence microscopy. The result from Figure 27A showed that GFP-INopt 3VI, M2 and M3 were both cytoplasmic and nuclear localized, in contrast to GFP-INopt wt, 3KR and M1 mutant which were exclusively localized to the nuclear compartment (Figure 27A). Nuclear localization of IN wt and 3KR is in line with the previous report that SUMO modification of IN at three Lys residues (K46/K136/K244) is not required for its nuclear translocation [214]. However, distinct localization pattern of IN putative SIMs mutants 3VI, M2 and M3 suggested the involvement of these two motifs (200IVDI203 and 257IKII260) in the nuclear translocation of IN protein.

To further confirm the impaired nuclear localization of IN 3VI mutant, 293T cells transfected with GFP-INopt wt or 3VI were collected and sequentially fractionated into cytoplasmic, nuclear and chromatin-bound extracts using Protein Fractionation Kit (Thermo Scientific) [115]. The presence of GFP-INopt wt and 3VI in these cellular compartments was detected by WB analysis (Figure 27B, left panel). The ratio of nuclear portion (a combination of Nuclear and Chromatin-bound from left panel) to cytoplasmic fractionate of proteins was calculated from their relative protein band intensities and shown as a pie chart in the right panel of Figure 27B. The nuclear localized portion of

GFP-INopt wt and 3VI in the total protein extracts is about 80% and 40%, respectively (Figure 27B, right panel). Taken together, results suggest that among the three putative SIMs of IN, at least 200IVDI203 and 257IKII260, are involved in its nuclear translocation by as yet unknown mechanisms.

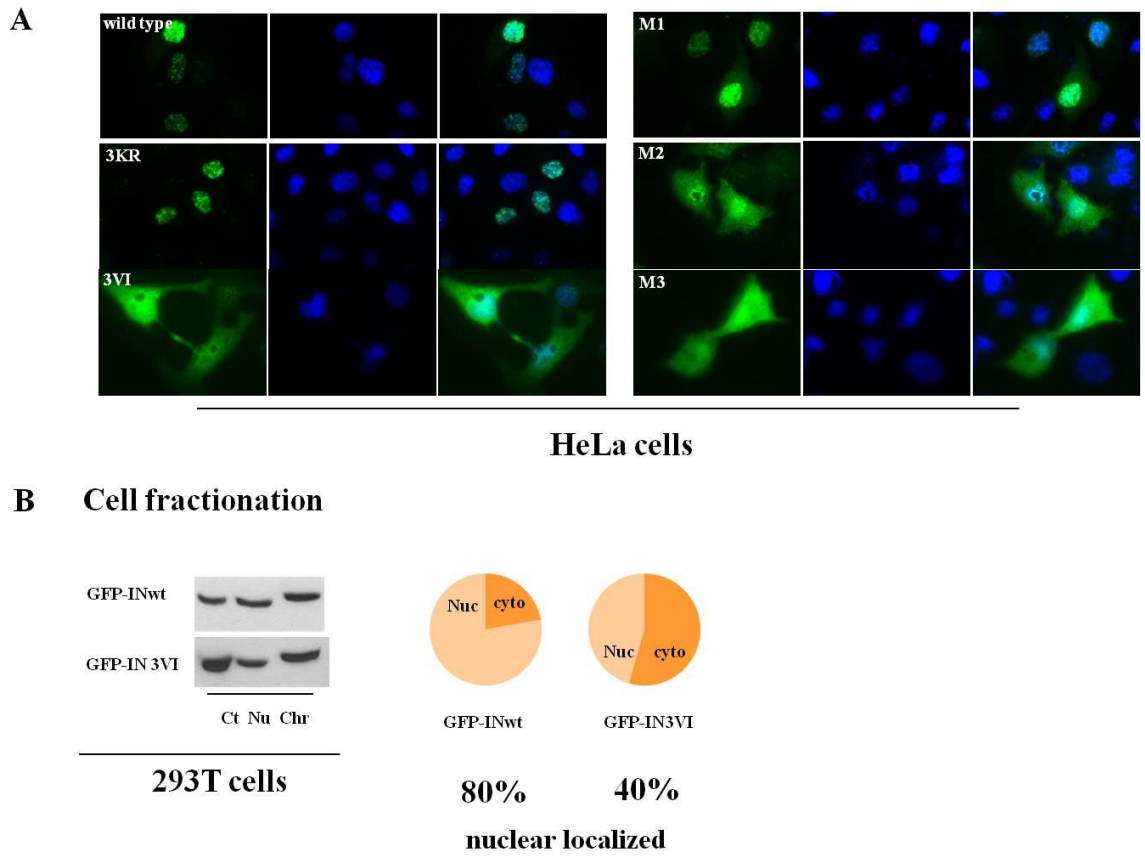

Figure 27 Effects of putative SIMs of IN on nuclear import of IN. The potential effects of IN SIMs in nuclear localization of IN were evaluated in HeLa cells (A) and 293T cells (B). A) HeLa cells were transfected with different GFP-INopt wt/mut. At 40 h post-transfection, cells were fixed, permeabilized and detected for GFP-INwt/mut localization by using anti-GFP antibody. The nuclei were stained with DAPI. B) 293T cells were transfected with GFP-INopt wt or 3VI for 48 h. The subcellular distribution patterns of IN wt and 3VI were determined by the subcellular fractionation method. Ct, cytoplasmic extracts; Nu, nuclear extracts; Chr, chromatin bound extract. The ratio of nuclear portion (a combination of Nu and Chr from left panel) to cytoplasmic fractionate of protein was calculated and shown as a pie chart in the right panel. The nuclear localized portion of GFP-INopt wt and 3VI in the total protein extracts is about 80% and

40%, respectively.

7.5 Discussion

SUMOylation of IN is newly identified PTM and the importance of SUMO modification of HIV-1 IN has not been well appreciated [214]. This ongoing study has explored the multiple roles of three putative HIV-1 IN SIMs including the regulation of IN SUMOylation, IN-LEDGF/p75 and IN-Ku70 interactions and nuclear localization of IN. Continuing studies of this project will aim to unravel whether these three putative SIMs are authentic SIMs sequence which can mediate non-covalent binding between HIV-1 IN and SUMO proteins, and to investigate the roles of these SIMs during HIV-1 replication.

7.5.1 SUMO binding and SUMO modification of HIV-1 IN

It has been proposed that non-covalent SUMO binding may facilitate the recognition and SUMOylation of SIM-containing proteins. In this model, SIM-SUMO interaction enhances the local concentration of SUMO E2 protein Ubc9 which has been attached to SUMO, thus allowing Ubc9 to be more efficient in catalyzing covalent binding of SUMO to the substrates. For instance, the interaction between SIM of Sp100 and SUMO-Ubc9 enhanced SUMOylation of Sp100 [349]. This SIM-dependent SUMOylation has also been described in various SUMO targets such as Daxx, RANBP2/Nup358, HIPK2 and BLM [347, 350, 351]. However, a few exceptions have been found in which SUMOylation of protein substrates inhibits SUMO binding or vice versa. For example, Srs2 SUMOylation inhibits its own SUMO binding, possibly due to reduced availability of its SIM motif for interaction with SUMOylated proteins in general by Srs2 SUMOylation [352]. Meanwhile, SUMOylation may potentially mask the protein-protein interaction site of the protein targets. For example, SUMOylation of the ubiquitin-

conjugating enzyme E2-25k could inhibit its interaction with the ubiquitin E1 enzyme, impairing the process of ubiquitination [243]. Thus, it appears that SUMO binding and SUMOylation can mutually regulate with each other, either positively or negatively.

In this study, we have generated the triple IN SIM mutant 3VI which simultaneously abolished all the potential SIMs in IN. Surprisingly, IN 3VI had four-fold increased SUMOylation level (Figure 24B). Therefore, this result suggested that potential SUMO binding of IN, at least these three IN sequences (72VILV75, 200IVDI203 and 257IKII260), negatively regulate its SUMOylation. Meanwhile, 3VI also had significantly elevated Ubc9 binding ability compared with IN wt in our study (Figure 24C). The role of SUMO E2 protein Ubc9 in the SUMO conjugation pathway, is to interact with the SUMO-E1 thioester complex and to accept SUMO then to bind target protein substrates, to which the SUMO molecule is attached [353]. Ubc9 is shown to directly bind SUMO conjugation consensus φKxE *in vitro* and in crystal studies [344, 345], and mutating ψKxE abolishes both Ubc9 binding and substrate modification [344]. Thus, Ubc9 binding to its target proteins is associated with SUMOylation of protein substrates. In our study, 3VI which harbors the putative SIMs in IN had increased Ubc9 binding affinity and SUMO3 modification, reiterating the notion that Ubc9 binding is a prerequisite for SUMO modification of substrates. This result thus suggested that putative SUMO binding of IN could block SUMO3 modification of HIV-1 IN. This could be explained by the possibility that SUMO binding of IN might block SUMOylation sites by recruiting SUMOylated binding partners or inducing conformational changes of IN, leading to the inaccessibilty of IN for SUMO modification.

3KR, a previously described IN mutant with mutations introduced to critical Lys residues within the three SUMO conjugation sites (45LKGE, 135IKQE and 243WKQE) [214], failed to perturb SUMOylation of IN by SUMO3 (Figure 24B). The discrepancy could be explained by different SUMO paralogues used in the present study. The earlier work done by Zamborlini and coworkers showed that 3KR had reduced SUMOylation level in the present of a mixture of three SUMO paralogues (SUMO1-3) in the *in vitro* SUMOylation assay while 3KR did not show obvious reduction of SUMO3 conjugation in the *in vivo* assay [214]. In our study, 3KR had similar SUMOylation level as wt when GFP-IN wt or 3KR was co-expressed with HA-SUMO3 proteins (Figure 24B). Meanwhile, 3KR still retained binding ability with Ubc9 which directly binds SUMO conjugation consensus φKxE and is required for all the SUMOs conjugation [344, 345]. Overall, our data suggest that these three Lys residues (K46/K136/K244) are not the major SUMO3 modification sties.

7.5.2 Consequence of SUMO binding of IN

The functional outcomes for non-covalent SUMO binding largely depend on SIM-SUMO interaction between SUMOylated proteins and SIM-containing binding partners. Thus, SUMO binding can promote or interfere interactions with proteins, DNA and other macromolecules at the molecular level, affecting protein stability, cytosolic-nuclear translocation, and transcriptional regulation (see a review [234]).

To investigate the potential effects of putative SIMs in IN in mediating binding with its SUMOylated cofactors, we have studied the interaction between HIV-1 IN with LEDGF/p75 and Ku70 both of which are modified by SUMO proteins and found that IN

putative SIMs mutant 3VI were severely impaired in T7-LEDGF/p75 binding while increasing the binding affinity with T7-Ku70 by 2-fold (Figure 25 and 26). However, 3VI bound Nup62 which is non-SUMOylated cofactor of IN [354], to the same level as wt (data not shown). These results together imply that these three putative SIMs of IN might be involved in the regulation of the interaction with its SUMO conjugated binding partners. We were interested in whether SIM-SUMO interaction could regulate IN-LEDGF/p75 binding. However, T7-LEDGF/p75 4K which were shown to be SUMOylation defective [346], could still bind IN as efficient as T7-LEDGF/p75 wt (Figure 25C). Since the direct binding of IN-LEDGF/p75 has also been revealed by cell-free *in vitro* binding assay [169, 172, 173], our data suggested that SUMO binding ability of IN might not directly affect its binding with LEDGF/p75. Rather, the disruption of these putative SIMs of IN might have profound impact on other functions of IN, indirectly influencing IN-LEDGF/p75 interaction. For example, V260 has been shown to be critical for multimerization of IN [355], and I200 is required for chromatin binding ability of IN [285]. With respect to the adverse impact of 3VI mutant on Ku70 binding, although Ku70 and LEDGF/p75 have different binding regions within IN (IN aa 230-288 for Ku70 and two binding regions for IN-LEDGF/p75 aa A128-W132 and I161- K173A) [154, 169, 171-173], we reasoned that IN 3VI mutant lost LEDGF/p75 binding ability and was rendered more accessible for binding other cofactors including Ku70, and components from SUMO pathways such as Ubc9 for more efficient SUMOylation.

SIM-SUMO interaction has a profound impact on the subcellular localization of SIM-containing proteins. For example, the SIM located at the C-terminal of Daxx mediated binding to SUMOylated promyelocytic leukaemia (PML) and localization to PML

nuclear bodies [347]. Another example is that SIMs of PML, Sp100 and hDaxx are required for the recruitment of these proteins to herpes simplex virus type 1 (HSV-1) induced foci, which also recruits SUMO proteins and SUMO E3 ligase PIAS2β [348]. In the present study, our results from both HeLa cells and 293T cells revealed that putative SIMs mutant IN 3VI were severely impaired for its nuclear localization (Figure 27A and 27B). Since SIM-containing proteins influence their subcellular localization often through SIM-SUMO interaction, it is conceivable that the impaired nuclear localization of IN 3VI might be caused by dysfunctional binding ability with its SUMO conjugated cofactors, the interaction which can drive IN to the nuclei. Such candidates expected to be SUMOylated and involved in the nuclear import of IN include LEDGF/p75, RANBP2/Nup258. Although contradictory findings have been reported, LEDGF/p75 was generally believed to function in nuclear import of IN [147, 165, 170]. RANBP2/Nup358 was first identified in genome-wide screens for cell factors functioning in the PIC nuclear import [116]. RanBP2 is a nucleoporin with SUMO E3 ligase activity that harbors both SIMs and SUMO conjugation sites [240, 350, 356, 357]. As SUMO E3 ligase, it has been shown to promote SUMOylation of a number of SUMO targets including Mdm2, HDAC4, topoisomerase II-alpha, PML or Sp100 [350, 358-360]. It is still unknown whether RANBP2 directly interacts with HIV-1 IN and functions in the nuclear import of IN. Hence, the impact of these putative SIMs on the nuclear import of IN still remains a mystery.

7.5.3 Ongoing works

The data presented so far have provided compelling evidence that three putative SIMs in IN negatively regulate SUMO E2 protein Ubc9 binding and SUMO3 modification, inversely affect the interactions with LEDGF/p75 and Ku70, also participate in the nuclear translocation of IN. The first three objectives of this project has been addressed in the result section and further investigation towards the last objective on assessing the SUMO binding ability of IN and its impacts on HIV-1 replication is being undertaken. We hypothesize that these putative SIMs (SIM1 72VILV75; SIM2 200IVDI203 and SIM3 257IKII260) of IN might mediate non-covalent binding with SUMO proteins. To test this hypothesis, we will investigate the interaction between IN wt or 3VI and SUMO proteins by *in vitro* and *in vivo* binding assay as described previously [351, 352]. In addition, further studies will also be required to study the potential effects of these putative SIMs in HIV-1 infection, more specifically, during the early stage of viral replication. To achieve this goal, various Vpr-RT-INwt/mut expresssors which will be used for packaging single-cycle replicating HIV-1 virus containing IN putative SIMs mutants were constructed. The same amount of IN wt/mut viruses (normalized by p24 value) will be used to infect C8166T cells. Viral infectivity of these viruses will be accessed by Luciferase assay at 48 h post-infection and their impact on different steps of the early stage of viral replication (including reverse transcription, nuclear import and integration) will be analyzed by quantitative real-time PCR as reported in earlier publications [215, 285]. Completion of these two key experiments will provide the answer to the questions of whether these three SIMs are authentic SIMs mediating non-covalent binding of SUMO proteins and how these SIMs of IN influence HIV-1 replication step(s).

Chapter 8 Major findings, general discussion and future works

8.1 Major findings

Although the availability of HAART has greatly improved health and life expectancy of HIV-1 infected patients, there are many challenges associated with current ARV drugs. Hence, there is an urgent need to develop new class anti-HIV drugs. Since HIV-1 replication heavily relies on host cellular machinery, the new strategy to develop cheaper and new drugs against HIV-1 is to block the key virus-host interactions. We are primarily focused in the study of cellular proteins and pathways important for retroviral infection, with a particular focus on the interplay between HIV-1 IN and its cellular cofactors. Better understanding on the host cellular proteins implicated in different aspects of IN will contribute to our knowledge of HIV-1 IN and viral replication, which will ultimately lead to the development of novel HIV-1 drugs.

The central hypothesis of this thesis is that HIV-1 IN utilizes various cellular cofactors to contribute to different aspects of IN and its related viral replication steps. In support of this central hypothesis, there were four primary goals. The first goal was to characterize the importance of chromatin binding and LEDGF/p75 binding abilities of HIV-1 IN in lethal phenotype induced by IN in yeast.

Based on the preliminary finding that IN mutants which failed to induce lethal phenotype in yeast cells lost its binding abilities with both host chromatin and cellular protein LEDGF/p75, the second goal was to map chromatin binding domain of IN within CCD and to address the question of whether HIV-1 IN can mediate host chromatin binding in LEDGF/p75-independent manner.

Thirdly, we aimed to investigate the interaction between IN and DNA repair protein Ku70 and to characterize the contribution of IN-Ku70 interaction to ubiquitin modification of IN as well as viral replication.

Lastly, our focus was to determine the roles of three putative SIMs of IN in its SUMO conjugation, cofactors binding and nuclear import, and to address the research question of whether these three putative SIMs indeed mediate non-covalent SUMO binding of IN and probe the functions of these sequences during HIV-1 replication.

There are two main arms for this project. The first arm (described in the first two objectives) was initiated from the study of lethal phenotype induced by IN in yeast cells and extend to investigate whether chromatin binding of IN is independent of IN-LEDGF/p75 interaction. The second arm (covered the last two objectives) focused on the cellular proteins-involved regulation of IN PTMs and their functions in the viral replication. To these ends, we combined molecular virology with yeast genetics, protein biochemistry and cell biology to accomplish these project objectives. The major findings from these objectives are summarized below.

8.1.1 Lethal phenotype induced by IN in yeast correlates with chromatin- and LEDGF/p75-binding abilities of HIV-1 IN

Yeast provides a simple but useful eukaryotic model for the study of many basic biological functions. The yeast expression system has been successfully adopted to investigate the various roles of IN including lethal phenotype induced by IN, nuclear import of IN and to identify IN-host factor interactions by yeast two-hybrid system [296, 361, 362]. The emergence of a lethal phenotype by the expression of HIV-1 IN in the

yeast *S. cerevisiae* was established earlier in our lab. And the preliminary data revealed that three IN mutants including V165A, A179P, KR186,7AA located in the C-terminal region of HIV-1 IN CCD failed to induce lethal phenotype. **In chapter 4**, our subsequent studies revealed that these IN mutants were impaired in their ability to associate with host chromatin and its cellular cofactor LEDGF/p75 but not Ini1, and that HIV-1 viruses harbouring these IN mutants were non-infectious. This work highlighted the importance of the C-terminal region of the HIV-1 IN CCD in its association with the chromatin of the host cell, viral replication and the IN-induced lethal phenotype in yeast. These data also support the previous observation that chromatin binding ability of IN is mediated by its interaction with host cofactor LEDGF/p75 or LEDGF/p75 serves as IN-to-chromatin tethering factor [170].

8.1.2 Uncoupled chromatin binding and LEDGF/p75 binding abilities of HIV-1 IN

The finding that three IN mutants V165A, A179P, KR186,7AA which were impaired for chromatin binding lost LEDGF/p75 binding affinity from **chapter 4** suggested that chromatin binding ability of IN might functionally correlate with LEDGF/p75 binding ability. Meanwhile, extensive studies have pinpointed LEDGF/p75 as the IN-to-chromatin tethering factor driving PICs to transcriptionally active regions of host chromosomes [150-156]. However, LEDGF/p75-independent chromatin targeting model, either through cofactor-independent or through alternative tethering factor mediated, was proposed to explain the low level of HIV-1 integration and replication in LEDGF/p75 depleted cells [120, 140, 176-178]. Thus, how IN alone, in the absence of LEDGF/p75, plays a role in chromatin binding still needs to be fully understood. **In chapter 5**, we performed site-directed mutagenic analysis at the C-terminal region of the IN CCD for

IN/chromatin binding and IN/LEDGF/p75 interaction. Results showed that several IN mutants including K159P, V165A, V176A, A179P, KR186,7AA and I203P were unable to bind both LEDGF/p75 and host chromatin. The mutants H171A, L172A and EH170,1AA, located in a loop region 170EHLK173 of IN, severely impaired their interaction with LEDGF/p75 but were still able to bind chromatin. Also, our data showed that LEDGF/p75 depletion in cells failed to dissociate IN from chromatin. Furthermore, the single-round HIV-1 replication assay results showed that the viruses harbouring IN mutants capable of LEDGF/p75-independent chromatin binding still sustained a low level of infection. However, failure to engage chromatin of IN is a replicative dead end for the virus. All of these data indicate that while LEDGF/p75 is important for productive HIV-1 replication, IN has the ability to bind chromatin in a LEDGF/p75-independent manner and sustain a low level HIV-1 infection.

8.1.3 Host protein Ku70 protects IN through two mechanisms

IN has been shown to be an unstable protein degraded by the N-end rule pathway through the host ubiquitin-proteasome machinery. However, it is still not fully understood how this viral protein is protected from the host UPS within cells during HIV replication. In the present study, we provide evidence that the host protein Ku70 interacts with HIV-1 IN and protects it from the Lys48-linked polyubiquitination proteasomal pathway. Moreover, Ku70 is able to down-regulate the overall protein polyubiquitination level within the host cells and to specifically deubiquitinate IN through their interaction. Mutagenic studies revealed that the C terminus of IN (aa 230–288) is required for binding to the N-terminal part of Ku70 (aa 1–430), and their interaction is independent of Ku70/80 heterodimerization. Finally, knockdown of Ku70 expression in both virus-

producing and target CD4+ T cells significantly disrupted HIV-1 replication and rendered 2-LTR circles and integration undetectable, indicating that Ku70 is required for both the early and the late stages of the HIV-1 life cycle. Interestingly, Ku70 was incorporated into the progeny virus in an IN-dependent way. We proposed that Ku70 may interact with IN during viral assembly and accompany HIV-1 IN upon entry into the new target cells, acting to 1) protect IN from the host defense system and 2) assist IN integration activity. Overall, this report provides another example of how HIV-1 manipulates host cellular machinery to protect the virus itself and to facilitate its replication.

8.1.4 Three putative SIMs of IN are required for LEDGF/p75 binding and its nuclear translocation

SUMOylation of IN is a newly identified PTM in 2011 and in contrast to other PTMs of IN such as ubiquitination and acetylation, the importance of SUMO modification of HIV-1 IN has not been appreciated [214]. SUMOylation affects various cellular processes, such as protein stability, localization, protein-protein interaction, DNA binding and activation (See a review [234]). Analysing the sequence of HIV-1 IN revealed that IN contains three putative SIMs V/I-V/I-x-V/I/L or V/I-x-V/I-V/I: SIM1 72VILV75; SIM2 200IVDI203 and SIM3 257IKII260. Further studies revealed that these three putative SIMs of IN negatively regulate its SUMO E2 protein Ubc9 binding and SUMO3 conjugation, and three previous identified SUMO conjugation sites (45LKGE, 135IKQE and 243WKQE) [214] were not required for SUMO3 modification in our *in vivo* SUMOylation assay. Mutations of these three putative SIMs exert different effects on IN-LEDGF/p75 and IN-Ku70 binding, showing severely weakened LEDGF/p75 binding but significantly increased binding affinity with Ku70. Furthermore, our study also showed

that mutations on the putative SIMs impaired nuclear localization of IN through a currently unknown mechanism. Thus, three important roles of these putative SIMs of IN including regulation of SUMOylation, cofactors binding and nuclear localization have been identified so far in this study.

8.2 General discussion and future work

8.2.1 The architecture of ternary complex IN/LEDGF/chromatin

Lethal phenotype induced by expression of IN in yeast cells has long been a puzzle. It has been ascribed to enzymatic activities of IN or yeast SNF5 gene which encodes a component of the SWI/SNF chromatin remodelling complex [295, 297, 301]. However, the contradictory finding suggested that catalytic activity of IN is not required for its induced lethal phenotype in yeast [298]. **In chapter 4**, we provided the evidence that lethal phenotype induced by IN in yeast is associated with its chromatin- and LEDGF/p75-binding abilities.

Extensive studies have been undertaken to investigate the interaction between IN and viral DNA by means of *in vitro* studies, and the mechanisms of their binding are relatively well understood (see a review [192]). However, the mechanism for chromatin recognition and binding of IN still remains elusive. To gain more insight into chromatin binding profile of IN, we carried out site-directed mutagenic analysis at the C-terminal region of the IN CCD responsible for IN-chromatin binding and IN-LEDGF/p75 interaction **in chapter 5**. Other than three previously identified IN mutants V165A, A179P, KR186,7AA, several IN mutants such as K159P, V176A, A179I, I203P were also severely impaired for host chromatin binding. The result from the chromatin binding

assay enabled the delineation of several important residues of IN essential for the formation of IN-chromatin complex.

The formation of this dynamic IN/LEDGF/chromatin ternary complex is fundamental for HIV-1 replication. It is well established that chromatin bound LEDGF/p75 tethers IN to chromatin [170, 363], and the presence of LEDGF/p75 promotes IN DNA-binding [364]. **In chapter 5**, we also discovered that IN mutants which lost LEDGF/p75 binding could still mediate chromatin binding within the loop region 170EHLK173 of IN. This result revealed previously undescribed, uncoupled chromatin- and LEDGF/p75-binding of IN, and it also provided structural details of IN/LEDGF/chromatin complex. After we published this research work, there are two independent papers describing the dynamic interplay among IN, LEDGF/p75 and chromatin. First, the presence of IN, in turn, was able to promote chromatin binding of LEDGF/p75 through their interaction [365]. Thus, it was proposed that IN and LEDGF/p75 act in concert with one another in their binding to chromatin: the binding of IN to LEDGF/p75 increases its affinity for DNA and subsequently IN enhances the binding of the LEDGF/p75-IN complex to chromatin. [365]. Another study investigated their relationship within the complex real-time in living HeLa cells [366]. It was shown that LEDGF/p75 does not permanently engage in the host chromatin; instead it moves about in nuclei of living cells in a chromatin hopping/scanning mode. This scan-and-lock model explaining the roles of LEDGF/p75 during IN-to-chromatin tethering suggests that the PWWP domain of LEDGF/p75 first locks the complex on chromatin, its IBD then interacts with HIV-1 IN, LEDGF/p75 lastly increases its affinity for chromatin [366].

Furthermore, our finding that IN mutants which lost LEDGF/p75 binding still bound host chromatin and supported low level HIV-1 replication opens up the intriguing question that whether HIV-1 could still integrate into host genome without the cellular tethering cofactor or there is an unidentified cellular partner involved in this process. In order to identify novel cellular cofactors of HIV-1 IN, the graduate student Kamal Uddin from our lab had isolated several IN's interacting proteins such as Hsp60, β-tubulin, γ-actin, ATP synthase alpha subunit and histone H1.2 by means of tandem affinity purification (TAP) system along with mass spectrometry [367]. This technique can be applied to identification of unknown chromatin associated cellular cofactor of HIV-1 IN from the chromatin bound fractionate isolated from LEDGF/p75 knockdown cells. With this technique, we hope to identify the potential candidates that can drive chromatin binding of IN when LEDGF/p75 is depleted.

8.2.2 Ku70 and LEDGF/p75: potential targets for anti-HIV therapy

Ku70 was identified as a new cellular cofactor of IN in this study and IN-Ku70 interaction was verified in overexpressed mammalian cells and HIV-1 infected target cells for the first time, as described **in chapter 6**. Ku70 has been shown to stabilize IN expression through down-regulating overall ubiquitin level in the cells and by reducing the ubiquitination of IN by specific IN-Ku70 interaction. Interestingly, cellular protein LEDGF/p75 is also able to increase IN expression and protects it from proteasome degradation through their interaction [175]. Our unpublished data showed that LEDGF/p75 and Ku70 indeed interact with each other in the cell-based Co-IP assay. To date, it is still unclear whether LEDGF/p75 and Ku70 act concertedly to protect IN from host proteasomal degradation pathway and the biological significance of LEDGF/p75-

Ku70 interaction during HIV-1 replication. Notably, **in chapter 7**, IN putative SIMs mutant 3VI almost abolished IN-LEDGF/p75 interaction while increasing Ku70 binding. Although the direct involvement of SUMO binding of IN in its binding affinity with LEDGF/p75 or Ku70 is presently under investigation, it implies that IN might have exclusive binding sites for LEDGF/p75 and Ku70 interactions. It is also necessary to define the organization of the protein architecture consisted of LEDGF/p75, Ku70 and IN, and to understand how they interact with and regulate each other and contribute to a more efficient integration. Future work will address these important questions.

In order to examine the role of Ku70 beyond shielding IN from proteasome degradation pathway in more detail, future work will also be directed toward understanding the possible role of Ku70 in the regulation of chromatin targeting of IN or integration site selection. The study done by Fink and colleagues showed that Ku heterodimer facilitates chromatin binding of the telomere binding protein TRF2 and that interaction with the chromatin may stabilize the TRF2 protein and prevent its proteasomal degradation [368]. Given that IN is a chromatin binding protein and is protected by Ku70 from proteasome degradation, it would be interesting to test the hypothesis that proteasomal protection of IN by host protein Ku70 would enhance chromatin association of IN. Additionally, it has been shown that heterozygous knockout and knockdown of Ku70 increased recombinant adeno-associated virus (rAAV) –mediated integration frequency in human somatic cells [369]. This finding not only has direct implication on rAAV-mediated knockout and gene therapy approaches, but also implies that Ku70, as part of the NHEJ pathway, might have a role in HIV-1 integration considering that it is incorporated into HIV-1 virion in the late

stage, and that it interacts with and prevents IN from host ubiquitous ubiquitin proteasome pathway during the early stage of viral replication.

Results from **chapter 6** showed that N-terminal part of Ku70 (Ku70 1-430) had stronger binding affinity to bind IN than Ku70 wt, and Ku70 knockdown in both virus-producing and target CD4+ T cells significantly disrupted HIV-1 replication. These data have provided proof-of-concept that IN-Ku70 interaction might represent novel, potentially druggable, anti-viral target. LEDGF/p75 is the most well understood cellular cofactor of IN and the interaction between LEDGF/p75 and IN has been extensively investigated as a new class of anti-IN target. It has been demonstrated that stable overexpression of the LEDGF/p75 IBD reduces HIV replication 100-fold and the replication block was mapped to the integration step [168, 178]. LEDGINs which are small molecule inhibitors that bind to the LEDGF/p75 binding pocket in HIV-1 IN have been proven to inhibit IN-LEDGF/p75 interaction and have potent antiviral activity [274]. The same basic strategies can be used when investigating IN-Ku70 interaction as antiviral target. For example, overexpression of Ku70 binding domain aa 1-430, or even smaller binding region, can potentially inhibit IN-Ku70 interaction and further block HIV-1 replication. And future solution structure of IN-Ku70 binding will not only shed light on the binding/interaction mechanism but also provide valuable information for structure-based drug design.

Two host DNA damage signaling and repair pathways including the NHEJ pathway (DNA-PK, Ku, Xrcc4, DNA ligase IV) and DNA damage-sensing (Atr (Atm and Rad related), γ-H2AX) pathways are suggested to participate in the successful gap repair during integration [203-205]. A latest study reported that LEDGF/p75 enhances DNA-

end resection and homologous recombination–mediated repair through its PWWP-dependent binding to chromatin [370]. It is still unknown currently whether HIV-1 recruit LEDGF/p75 to enhance homologous DNA repair when infected cells enter S phase with retroviral DNA intermediates unrepaired. NHEJ is known to repair DNA double strand break throughout all the cell cycle phases mainly occurring in the G1 phase, while homologous recombination which repairs DNA break with high fidelity predominates the repair during the S phase [371]. Therefore, it is tempting to speculate that HIV-1 utilizes the complementary Ku70 and LEDGF/p75 to be corporately and actively involved in host proteins-mediated gap repair phase of HIV-1 integration when the cells are either in G1 phase or S phase. Undoubtedly, the future will see more work in this important process implicated with these two multifaceted cellular proteins.

8.2.3 SUMOylation and SUMO binding of IN

The only study that described SUMO modification of IN identified three ψ-K-x-D/E SUMOylation sites (45LKGE, 135IKQE and 243WKQE) within IN [214]. However, it still remains elusive as to which SUMO subtype(s) preferentially modifies IN. Interestingly, these three SUMOylation sites are not SUMO3 modification sites as revealed by *in vivo* SUMOylation assay where IN 3KR which has all key Lys residues mutated into Arg still retains the same level of SUMOylation level as IN wt. In line with this result, 3KR has the similar binding affinity with SUMO E2 conjugation enzyme Ubc9 when compared with IN wt, which is in agreement with unaltered SUMO3 modification in IN 3KR mutant.

In addition to SUMO covalent modification, we were also interested in non-covalent SUMO binding of IN and its roles involved. Three putative SIMs of IN (SIM1 72VILV75; SIM2 200IVDI203 and SIM3 257IKII260) which fit into the SIM consensus V/I-x-V/I-V/I or V/I-V/I-x-V/I/L (where x can be any amino acid) were examined in this study. **In chapter 7**, these putative SIMs of IN were shown to negatively regulate SUMOylation of IN by SUMO3, differently influence on LEDGF/p75 and Ku70 binding and participate in nuclear import of IN. The results showed that IN 3VI mutant which had hydrophobic Val and Ile in 72VI73, 200IV201 and 259II260 mutated into Ala largely diminished LEDGF/p75 binding and moderately increased Ku70 binding, while bound non-SUMOylated cofactor Nup62 to the same extent as IN wt. Thus, we speculate that the three putative SIMs of IN might be involved in the regulation of the interaction with its SUMO conjugated binding partners. Another interesting observation is that 3VI, M2 and M3 mutants were both cytoplasmic and nuclear distributed, in contrast to exclusively nuclear localization of IN wt. The roles of these putative SIMs in the nuclear localization of IN still await further characterization. Since 3VI, M2 and M3 all had decreased LEDGF/p75 binding, it raised the possibility that nuclear redistribution of these mutants might be caused by impaired IN-LEDGF/p75 binding [147, 165, 170]. However, we cannot rule out the possibility that other cellular proteins such as RANBP2/Nup358 might contribute to this process. In support of this hypothesis, RANBP2/Nup358 was involved in the PIC nuclear import , and it is a SUMO E3 ligase that harbors both SIMs and SUMO conjugation sites [116, 240, 350, 356, 357]. Despite all these critical roles that these putative SIMs play, the exact mechanisms still remain elusive.

Current and future work in this project will focus on the following key questions: 1) whether these three putative SIMs or which specific SIM of IN could mediate non-covalent SUMO binding of IN; 2) whether and how these important motifs affect HIV-1 viral replication. Yeast two-hybird assay and *in vitro* binding assay are the two main approaches being employed to investigate non-covalent SUMO binding of certain proteins [218, 372]. The *in vitro* binding assay studies the binding between recombinant epitope-tagged SUMO proteins and proteins of interest. To address the first question, GST pull down assay will be performed to study the potential interaction between recombinant IN wt and recombinant GST-SUMO1, GST-SUMO2 or GST-SUMO3 fusion proteins. Then we will express recombinant IN 3VI proteins in bacteria and analyse its binding with SUMO fusion protein. GST alone will be included as a negative control. We speculate that IN wt is able to bind SUMO in the GST pull down assay whereas IN 3VI might lose SUMO binding ability. Though this assay, we will be able to demonstrate whether these three putative SIMs are authentic SIMs that mediate non-covalent SUMO binding of IN.

To investigate the functional roles of these important sequences during HIV-1 replication, single cycle replicating HIV-1 virus harbouring IN mutants M1, M2 and M3 will be generated by packaging VSV-G, Vpr-RT-IN and RT/IN/Env gene-deleted NLlucΔBglΔRI provirus into 293T cells, as described earlier [215, 285]. These viruses containing various IN mutants will be used to infect C8166 CD4+T cells and viral infectivity will be assessed by Luciferase assay at 48 h post-infection. To pinpoint which early step of viral replication might be blocked in these single cycle replicating viruses, real-time PCR will be carried out to quantify late RT, 2-LTR circles and integrated DNA

which are indicative of reverse transcription, nuclear import of PIC and integration in the viral life cycle [215, 285]. Although it is currently unknown the effects of these IN mutants on HIV-1 replication, it is expected that these viruses, at least for the viruses containing M2 and M3 mutants which were impaired for LEDGF/p75 binding and nuclear import of IN, are replication and integration defective. Future studies aimed at investigating the detailed biological action of these IN putative SIMs mutant viruses will provide important clues to the understanding of the involvement of SUMO binding of IN in the retroviral replication cycle.

8.3 Potential applications

The first FDA-approved HIV IN inhibitor RAL blocks integration by binding to catalytic metal cations in the catalytic core and preventing covalent bonds forming between IN and host DNA [373]. Current research effort is focused on virus-host protein/DNA interactions which are required for the completion of integration and the HIV replication cycle. The development of new IN inhibitors will be driven by an improved understanding of IN/cofactors/chromatin structure and of cellular cofactors-regulated pre-integration and integration events. **In chapter 4 and 5**, we characterized two critical molecular interactions that determine the fate of HIV-1 integration and viral replication: IN with host chromatin and IN with the cellular cofactor LEDGF/p75. **Chapter 6** described another cellular cofactor Ku70 and its roles in the ubiquitin modification of IN. Lastly, **chapter 7** described the roles of putative IN SIMs in SUMO modification, cofactors binding and nuclear localization. Our hope is to translate this basic knowledge into novel targets for antiviral therapies, by developing inhibitors of host-virus protein interactions.

8.4 Concluding remarks

A hallmark of retroviral replication is the permanent integration of the viral genome into the host cell genome, which is catalyzed by the viral enzyme IN. The process of integration is accompanied by the dynamic interactions among IN and viral DNA, chromatin, host proteins within the multiprotein complex PIC. Our goal was to determine the molecular mechanisms of cellular cofactors involved regulation of HIV-1 integration, with a particular emphasis on the interplay between IN with chromatin and cellular cofactors LEDGF/p75, Ku70. We have shown that 1) lethal phenotype induced by HIV-1 IN in yeast might be related to its capability of chromatin association and LEDGF/p75 binding; 2) HIV-1 IN is able to mediate chromatin binding in LEDGF/p75-independent manner; 3) IN is protected by host DNA repair protein Ku70 from proteasomal degradation pathway; 4) IN harbors three putative SIMs which have been shown to affect SUMO modification, cofactors binding and nuclear localization of IN. Overall, these results obtained from this thesis work highlighted the roles of cellular cofactors of HIV-1 IN (including LEDGF/p75 and Ku70) in the regulation of various actions of IN and viral replication.

Chapter 9 Reference

1. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, Mann D, Sidhu GD, Stahl RE, Zolla-Pazner S, et al: **Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS).** *Science* 1983, **220:**865-867.
2. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, et al: **Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).** *Science* 1983, **220:**868-871.
3. UNAIDS: **World AIDS Day report 2011.** 2011.
4. McCutchan FE: **Global epidemiology of HIV.** *J Med Virol* 2006, **78 Suppl 1:**S7-S12.
5. McCutchan FE: **Understanding the genetic diversity of HIV-1.** *Aids* 2000, **14 Suppl 3:**S31-44.
6. Zhu P, Liu J, Bess J, Jr., Chertova E, Lifson JD, Grise H, Ofek GA, Taylor KA, Roux KH: **Distribution and three-dimensional structure of AIDS virus envelope spikes.** *Nature* 2006, **441:**847-852.
7. Turner BG, Summers MF: **Structural biology of HIV.** *J Mol Biol* 1999, **285:**1-32.
8. Giroud C, Chazal N, Briant L: **Cellular kinases incorporated into HIV-1 particles: passive or active passengers?** *Retrovirology* 2011, **8:**71.
9. Ratner L, Haseltine W, Patarca R, Livak KJ, Starcich B, Josephs SF, Doran ER, Rafalski JA, Whitehorn EA, Baumeister K, et al.: **Complete nucleotide sequence of the AIDS virus, HTLV-III.** *Nature* 1985, **313:**277-284.
10. Wain-Hobson S: **HIV genome variability in vivo.** *Aids* 1989, **3 Suppl 1:**S13-18.
11. Frankel AD, Young JA: **HIV-1: fifteen proteins and an RNA.** *Annu Rev Biochem* 1998, **67:**1-25.
12. Goff SP: **Host factors exploited by retroviruses.** *Nat Rev Microbiol* 2007, **5:**253-263.
13. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, Burch CL, Weeks KM: **Architecture and secondary structure of an entire HIV-1 RNA genome.** *Nature* 2009, **460:**711-716.
14. Malim MH, Emerman M: **HIV-1 Accessory Proteins Ensuring Viral Survival in a Hostile Environment.** *Cell host & microbe* 2008, **3:**388-398.
15. Gramberg T, Sunseri N, Landau NR: **Accessories to the crime: recent advances in HIV accessory protein biology.** *Curr HIV/AIDS Rep* 2009, **6:**36-42.
16. Gallay P, Hope T, Chin D, Trono D: **HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway.** *Proc Natl Acad Sci U S A* 1997, **94:**9825-9830.
17. Gottlinger HG, Sodroski JG, Haseltine WA: **Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1.** *Proc Natl Acad Sci U S A* 1989, **86:**5781-5785.
18. Ganser-Pornillos BK, Yeager M, Sundquist WI: **The structural biology of HIV assembly.** *Curr Opin Struct Biol* 2008, **18:**203-217.

19. Miller MD, Farnet CM, Bushman FD: **Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition.** *J Virol* 1997, **71:**5382-5390.
20. Bukrinskaya AG, Ghorpade A, Heinzinger NK, Smithgall TE, Lewis RE, Stevenson M: **Phosphorylation-dependent human immunodeficiency virus type 1 infection and nuclear targeting of viral DNA.** *Proc Natl Acad Sci U S A* 1996, **93:**367-371.
21. Haffar O, Bukrinsky M: **Nuclear translocation as a novel target for anti-HIV drugs.** *Expert Rev Anti Infect Ther* 2005, **3:**41-50.
22. Fiorentini S, Marini E, Caracciolo S, Caruso A: **Functions of the HIV-1 matrix protein p17.** *New Microbiol* 2006, **29:**1-10.
23. Fiorentini S, Riboldi E, Facchetti F, Avolio M, Fabbri M, Tosti G, Becker PD, Guzman CA, Sozzani S, Caruso A: **HIV-1 matrix protein p17 induces human plasmacytoid dendritic cells to acquire a migratory immature cell phenotype.** *Proc Natl Acad Sci U S A* 2008, **105:**3867-3872.
24. Renga B, Francisci D, D'Amore C, Schiaroli E, Mencarelli A, Cipriani S, Baldelli F, Fiorucci S: **The HIV Matrix Protein p17 Subverts Nuclear Receptors Expression and Induces a STAT1-Dependent Proinflammatory Phenotype in Monocytes.** *PLoS One* 2012, **7:**e35924.
25. Giagulli C, Magiera AK, Bugatti A, Caccuri F, Marsico S, Rusnati M, Vermi W, Fiorentini S, Caruso A: **HIV-1 matrix protein p17 binds to the IL-8 receptor CXCR1 and shows IL-8-like chemokine activity on monocytes through Rho/ROCK activation.** *Blood* 2012, **119:**2274-2283.
26. Stremlau M, Perron M, Lee M, Li Y, Song B, Javanbakht H, Diaz-Griffero F, Anderson DJ, Sundquist WI, Sodroski J: **Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor.** *Proc Natl Acad Sci U S A* 2006, **103:**5514-5519.
27. Li F, Goila-Gaur R, Salzwedel K, Kilgore NR, Reddick M, Matallana C, Castillo A, Zoumplis D, Martin DE, Orenstein JM, et al: **PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing.** *Proc Natl Acad Sci U S A* 2003, **100:**13555-13560.
28. Kelly BN, Kyere S, Kinde I, Tang C, Howard BR, Robinson H, Sundquist WI, Summers MF, Hill CP: **Structure of the antiviral assembly inhibitor CAP-1 complex with the HIV-1 CA protein.** *J Mol Biol* 2007, **373:**355-366.
29. Thomas JA, Gorelick RJ: **Nucleocapsid protein function in early infection processes.** *Virus Res* 2008, **134:**39-63.
30. Barat C, Schatz O, Le Grice S, Darlix JL: **Analysis of the interactions of HIV1 replication primer tRNA(Lys,3) with nucleocapsid protein and reverse transcriptase.** *J Mol Biol* 1993, **231:**185-190.
31. Wu W, Henderson LE, Copeland TD, Gorelick RJ, Bosche WJ, Rein A, Levin JG: **Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse transcriptase pausing at a secondary structure near the murine leukemia virus polypurine tract.** *J Virol* 1996, **70:**7132-7142.
32. Lapadat-Tapolsky M, De Rocquigny H, Van Gent D, Roques B, Plasterk R, Darlix JL: **Interactions between HIV-1 nucleocapsid protein and viral DNA**

may have important functions in the viral life cycle. *Nucleic Acids Res* 1993, **21:**831-839.
33. Fossen T, Wray V, Bruns K, Rachmat J, Henklein P, Tessmer U, Maczurek A, Klinger P, Schubert U: **Solution structure of the human immunodeficiency virus type 1 p6 protein.** *J Biol Chem* 2005, **280:**42515-42527.
34. Beilhartz GL, Gotte M: **HIV-1 Ribonuclease H: Structure, Catalytic Mechanism and Inhibitors.** *Viruses* 2010, **2:**900-926.
35. Navia MA, Fitzgerald PM, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP: **Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1.** *Nature* 1989, **337:**615-620.
36. Wensing AM, van Maarseveen NM, Nijhuis M: **Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance.** *Antiviral Res* 2010, **85:**59-74.
37. Davenport TM, Friend D, Ellingson K, Xu H, Caldwell Z, Sellhorn G, Kraft Z, Strong RK, Stamatatos L: **Binding interactions between soluble HIV envelope glycoproteins and quaternary-structure-specific monoclonal antibodies PG9 and PG16.** *J Virol* 2011, **85:**7095-7107.
38. Li L, Li HS, Pauza CD, Bukrinsky M, Zhao RY: **Roles of HIV-1 auxiliary proteins in viral pathogenesis and host-pathogen interactions.** *Cell Res* 2005, **15:**923-934.
39. Debaisieux S, Rayne F, Yezid H, Beaumelle B: **The ins and outs of HIV-1 Tat.** *Traffic* 2012, **13:**355-363.
40. Levin A, Rosenbluh J, Hayouka Z, Friedler A, Loyter A: **Integration of HIV-1 DNA is regulated by interplay between viral rev and cellular LEDGF/p75 proteins.** *Mol Med* 2010, **16:**34-44.
41. Malim MH, Emerman M: **HIV-1 accessory proteins--ensuring viral survival in a hostile environment.** *Cell Host Microbe* 2008, **3:**388-398.
42. Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, Yu XF: **Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex.** *Science* 2003, **302:**1056-1060.
43. Marin M, Rose KM, Kozak SL, Kabat D: **HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation.** *Nat Med* 2003, **9:**1398-1403.
44. Yung E, Sorin M, Wang EJ, Perumal S, Ott D, Kalpana GV: **Specificity of interaction of INI1/hSNF5 with retroviral integrases and its functional significance.** *J Virol* 2004, **78:**2222-2231.
45. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L: **The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA.** *Nature* 2003, **424:**94-98.
46. Nascimbeni M, Bouyac M, Rey F, Spire B, Clavel F: **The replicative impairment of Vif- mutants of human immunodeficiency virus type 1 correlates with an overall defect in viral DNA synthesis.** *J Gen Virol* 1998, **79 (Pt 8):**1945-1950.
47. Kotler M, Simm M, Zhao YS, Sova P, Chao W, Ohnona SF, Roller R, Krachmarov C, Potash MJ, Volsky DJ: **Human immunodeficiency virus type 1 (HIV-1) protein Vif inhibits the activity of HIV-1 protease in bacteria and in vitro.** *J Virol* 1997, **71:**5774-5781.

48. Akari H, Fujita M, Kao S, Khan MA, Shehu-Xhilaga M, Adachi A, Strebel K: **High level expression of human immunodeficiency virus type-1 Vif inhibits viral infectivity by modulating proteolytic processing of the Gag precursor at the p2/nucleocapsid processing site.** *J Biol Chem* 2004, **279:**12355-12362.
49. Strebel K, Klimkait T, Maldarelli F, Martin MA: **Molecular and biochemical analyses of human immunodeficiency virus type 1 vpu protein.** *J Virol* 1989, **63:**3784-3791.
50. Akari H, Bour S, Kao S, Adachi A, Strebel K: **The human immunodeficiency virus type 1 accessory protein Vpu induces apoptosis by suppressing the nuclear factor kappaB-dependent expression of antiapoptotic factors.** *J Exp Med* 2001, **194:**1299-1311.
51. Klimkait T, Strebel K, Hoggan MD, Martin MA, Orenstein JM: **The human immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release.** *J Virol* 1990, **64:**621-629.
52. Willey RL, Maldarelli F, Martin MA, Strebel K: **Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4.** *J Virol* 1992, **66:**7193-7200.
53. Deacon NJ, Tsykin A, Solomon A, Smith K, Ludford-Menting M, Hooker DJ, McPhee DA, Greenway AL, Ellett A, Chatfield C, et al: **Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients.** *Science* 1995, **270:**988-991.
54. Kirchhoff F, Greenough TC, Brettler DB, Sullivan JL, Desrosiers RC: **Brief report: absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection.** *N Engl J Med* 1995, **332:**228-232.
55. Foster JL, Garcia JV: **Role of Nef in HIV-1 replication and pathogenesis.** *Adv Pharmacol* 2007, **55:**389-409.
56. Coiras M, Lopez-Huertas MR, Sanchez del Cojo M, Mateos E, Alcami J: **Dual role of host cell factors in HIV-1 replication: restriction and enhancement of the viral cycle.** *AIDS Rev* 2010, **12:**103-112.
57. Bushman FD, Malani N, Fernandes J, D'Orso I, Cagney G, Diamond TL, Zhou H, Hazuda DJ, Espeseth AS, Konig R, et al: **Host cell factors in HIV replication: meta-analysis of genome-wide studies.** *PLoS Pathog* 2009, **5:**e1000437.
58. Nisole S, Saib A: **Early steps of retrovirus replicative cycle.** *Retrovirology* 2004, **1:**9.
59. Gallo SA, Finnegan CM, Viard M, Raviv Y, Dimitrov A, Rawat SS, Puri A, Durell S, Blumenthal R: **The HIV Env-mediated fusion reaction.** *Biochim Biophys Acta* 2003, **1614:**36-50.
60. Auewarakul P, Wacharapornin P, Srichatrapimuk S, Chutipongtanate S, Puthavathana P: **Uncoating of HIV-1 requires cellular activation.** *Virology* 2005, **337:**93-101.
61. Misumi S, Inoue M, Dochi T, Kishimoto N, Hasegawa N, Takamune N, Shoji S: **Uncoating of human immunodeficiency virus type 1 requires prolyl isomerase Pin1.** *J Biol Chem* 2010, **285:**25185-25195.
62. Abbondanzieri EA, Bokinsky G, Rausch JW, Zhang JX, Le Grice SF, Zhuang X: **Dynamic binding orientations direct activity of HIV reverse transcriptase.** *Nature* 2008, **453:**184-189.

63. Zhang H, Zhang Y, Spicer TP, Abbott LZ, Abbott M, Poiesz BJ: **Reverse transcription takes place within extracellular HIV-1 virions: potential biological significance.** *AIDS Res Hum Retroviruses* 1993, **9:**1287-1296.
64. Warrilow D, Tachedjian G, Harrich D: **Maturation of the HIV reverse transcription complex: putting the jigsaw together.** *Rev Med Virol* 2009, **19:**324-337.
65. Farnet CM, Haseltine WA: **Integration of human immunodeficiency virus type 1 DNA in vitro.** *Proc Natl Acad Sci U S A* 1990, **87:**4164-4168.
66. Ellison V, Abrams H, Roe T, Lifson J, Brown P: **Human immunodeficiency virus integration in a cell-free system.** *J Virol* 1990, **64:**2711-2715.
67. Al-Mawsawi LQ, Neamati N: **Blocking interactions between HIV-1 integrase and cellular cofactors: an emerging anti-retroviral strategy.** *Trends Pharmacol Sci* 2007, **28:**526-535.
68. Van Maele B, Busschots K, Vandekerckhove L, Christ F, Debyser Z: **Cellular co-factors of HIV-1 integration.** *Trends Biochem Sci* 2006, **31:**98-105.
69. Freed EO: **HIV-1 replication.** *Somat Cell Mol Genet* 2001, **26:**13-33.
70. Wu Y, Marsh JW: **Gene transcription in HIV infection.** *Microbes Infect* 2003, **5:**1023-1027.
71. Gale M, Jr., Tan SL, Katze MG: **Translational control of viral gene expression in eukaryotes.** *Microbiol Mol Biol Rev* 2000, **64:**239-280.
72. Gheysen D, Jacobs E, de Foresta F, Thiriart C, Francotte M, Thines D, De Wilde M: **Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells.** *Cell* 1989, **59:**103-112.
73. Bryant M, Ratner L: **Myristoylation-dependent replication and assembly of human immunodeficiency virus 1.** *Proc Natl Acad Sci U S A* 1990, **87:**523-527.
74. Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI: **Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly.** *Proc Natl Acad Sci U S A* 1996, **93:**3099-3104.
75. Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, Klein CA, Swanstrom R: **The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions.** *J Virol* 1994, **68:**8017-8027.
76. Zheng R, Jenkins TM, Craigie R: **Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity.** *Proc Natl Acad Sci U S A* 1996, **93:**13659-13664.
77. Bushman FD, Engelman A, Palmer I, Wingfield P, Craigie R: **Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding.** *Proc Natl Acad Sci U S A* 1993, **90:**3428-3432.
78. Lutzke RA, Plasterk RH: **Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding.** *J Virol* 1998, **72:**4841-4848.
79. Berthoux L, Sebastian S, Muesing MA, Luban J: **The role of lysine 186 in HIV-1 integrase multimerization.** *Virology* 2007, **364:**227-236.

80. Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR: **Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium.** *Proc Natl Acad Sci U S A* 1998, **95:**9150-9154.
81. Turlure F, Devroe E, Silver PA, Engelman A: **Human cell proteins and human immunodeficiency virus DNA integration.** *Front Biosci* 2004, **9:**3187-3208.
82. Mousnier A, Kubat N, Massias-Simon A, Segeral E, Rain JC, Benarous R, Emiliani S, Dargemont C: **von Hippel Lindau binding protein 1-mediated degradation of integrase affects HIV-1 gene expression at a postintegration step.** *Proc Natl Acad Sci U S A* 2007, **104:**13615-13620.
83. Woerner AM, Klutch M, Levin JG, Marcus-Sekura CJ: **Localization of DNA binding activity of HIV-1 integrase to the C-terminal half of the protein.** *AIDS Res Hum Retroviruses* 1992, **8:**297-304.
84. Lutzke RA, Vink C, Plasterk RH: **Characterization of the minimal DNA-binding domain of the HIV integrase protein.** *Nucleic Acids Res* 1994, **22:**4125-4131.
85. Titolo S, Brault K, Majewski J, White PW, Archambault J: **Characterization of the minimal DNA binding domain of the human papillomavirus e1 helicase: fluorescence anisotropy studies and characterization of a dimerization-defective mutant protein.** *J Virol* 2003, **77:**5178-5191.
86. Ao Z, Danappa Jayappa K, Wang B, Zheng Y, Kung S, Rassart E, Depping R, Kohler M, Cohen EA, Yao X: **Importin alpha3 interacts with HIV-1 integrase and contributes to HIV-1 nuclear import and replication.** *J Virol* 2010, **84:**8650-8663.
87. Wilkinson TA, Januszyk K, Phillips ML, Tekeste SS, Zhang M, Miller JT, Le Grice SF, Clubb RT, Chow SA: **Identifying and characterizing a functional HIV-1 reverse transcriptase-binding site on integrase.** *J Biol Chem* 2009, **284:**7931-7939.
88. Zhu K, Dobard C, Chow SA: **Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase.** *J Virol* 2004, **78:**5045-5055.
89. McKee CJ, Kessl JJ, Shkriabai N, Dar MJ, Engelman A, Kvaratskhelia M: **Dynamic modulation of HIV-1 integrase structure and function by cellular lens epithelium-derived growth factor (LEDGF) protein.** *J Biol Chem* 2008, **283:**31802-31812.
90. Engelman A, Englund G, Orenstein JM, Martin MA, Craigie R: **Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication.** *J Virol* 1995, **69:**2729-2736.
91. Tsurutani N, Kubo M, Maeda Y, Ohashi T, Yamamoto N, Kannagi M, Masuda T: **Identification of critical amino acid residues in human immunodeficiency virus type 1 IN required for efficient proviral DNA formation at steps prior to integration in dividing and nondividing cells.** *J Virol* 2000, **74:**4795-4806.
92. Wu X, Liu H, Xiao H, Conway JA, Hehl E, Kalpana GV, Prasad V, Kappes JC: **Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions with the nucleoprotein reverse transcription complex.** *J Virol* 1999, **73:**2126-2135.

93. Leavitt AD, Robles G, Alesandro N, Varmus HE: **Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection.** *J Virol* 1996, **70:**721-728.
94. Masuda T, Planelles V, Krogstad P, Chen IS: **Genetic analysis of human immunodeficiency virus type 1 integrase and the U3 att site: unusual phenotype of mutants in the zinc finger-like domain.** *J Virol* 1995, **69:**6687-6696.
95. Hehl EA, Joshi P, Kalpana GV, Prasad VR: **Interaction between human immunodeficiency virus type 1 reverse transcriptase and integrase proteins.** *J Virol* 2004, **78:**5056-5067.
96. Nishitsuji H, Hayashi T, Takahashi T, Miyano M, Kannagi M, Masuda T: **Augmentation of reverse transcription by integrase through an interaction with host factor, SIP1/Gemin2 Is critical for HIV-1 infection.** *PLoS One* 2009, **4:**e7825.
97. Warren K, Warrilow D, Meredith L, Harrich D: **Reverse Transcriptase and Cellular Factors: Regulators of HIV-1 Reverse Transcription.** *Viruses* 2009, **1:**873-894.
98. Wang JY, Ling H, Yang W, Craigie R: **Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein.** *Embo J* 2001, **20:**7333-7343.
99. Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, Leavitt AD, Stroud RM: **Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding.** *Proc Natl Acad Sci U S A* 2000, **97:**8233-8238.
100. Yamashita M, Emerman M: **Retroviral infection of non-dividing cells: old and new perspectives.** *Virology* 2006, **344:**88-93.
101. Suzuki Y, Craigie R: **The road to chromatin - nuclear entry of retroviruses.** *Nat Rev Microbiol* 2007, **5:**187-196.
102. Lewis P, Hensel M, Emerman M: **Human immunodeficiency virus infection of cells arrested in the cell cycle.** *Embo J* 1992, **11:**3053-3058.
103. Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, Lifson JD, Bonhoeffer S, Nowak MA, Hahn BH, et al.: **Viral dynamics in human immunodeficiency virus type 1 infection.** *Nature* 1995, **373:**117-122.
104. Karni O, Friedler A, Zakai N, Gilon C, Loyter A: **A peptide derived from the N-terminal region of HIV-1 Vpr promotes nuclear import in permeabilized cells: elucidation of the NLS region of the Vpr.** *FEBS Lett* 1998, **429:**421-425.
105. Bouyac-Bertoia M, Dvorin JD, Fouchier RA, Jenkins Y, Meyer BE, Wu LI, Emerman M, Malim MH: **HIV-1 infection requires a functional integrase NLS.** *Mol Cell* 2001, **7:**1025-1035.
106. Fouchier RAM, Meyer BE, Simon JHM, Fischer U, Malim MH: **HIV-1 infection of non-dividing cells: evidence that the amino-terminal basic region of the viral matrix protein is important for Gag processing but not for post-entry nuclear import.** *EMBO J* 1997, **16:**4531-4539.
107. Yamashita M, Emerman M: **The cell cycle independence of HIV infections is not determined by known karyophilic viral elements.** *PLoS Pathog* 2005, **1:**e18.

108. Seamon JA, Jones KS, Miller C, Roth MJ: **Inserting a nuclear targeting signal into a replication-competent Moloney murine leukemia virus affects viral export and is not sufficient for cell cycle-independent infection.** *J Virol* 2002, **76:**8475-8484.
109. Ao Z, Huang G, Yao H, Xu Z, Labine M, Cochrane AW, Yao X: **Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication.** *J Biol Chem* 2007, **282:**13456-13467.
110. Jayappa KD, Ao Z, Yang M, Wang J, Yao X: **Identification of critical motifs within HIV-1 integrase required for importin alpha3 interaction and viral cDNA nuclear import.** *J Mol Biol* 2011, **410:**847-862.
111. Petit C, Schwartz O, Mammano F: **The karyophilic properties of human immunodeficiency virus type 1 integrase are not required for nuclear import of proviral DNA.** *J Virol* 2000, **74:**7119-7126.
112. Lu R, Limon A, Devroe E, Silver PA, Cherepanov P, Engelman A: **Class II integrase mutants with changes in putative nuclear localization signals are primarily blocked at a postnuclear entry step of human immunodeficiency virus type 1 replication.** *J Virol* 2004, **78:**12735-12746.
113. Jayappa KD, Ao Z, Yao X: **The HIV-1 passage from cytoplasm to nucleus: the process involving a complex exchange between the components of HIV-1 and cellular machinery to access nucleus and successful integration.** *Int J Biochem Mol Biol* 2012, **3:**70-85.
114. Christ F, Thys W, De Rijck J, Gijsbers R, Albanese A, Arosio D, Emiliani S, Rain JC, Benarous R, Cereseto A, Debyser Z: **Transportin-SR2 imports HIV into the nucleus.** *Curr Biol* 2008, **18:**1192-1202.
115. Ao Z, Jayappa KD, Wang B, Zheng Y, Wang X, Peng J, Yao X: **Contribution of host nucleoporin 62 in HIV-1 integrase chromatin association and viral DNA integration.** *J Biol Chem* 2012, **287:**10544-10555.
116. Konig R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Irelan JT, Chiang CY, Tu BP, De Jesus PD, Lilley CE, et al: **Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication.** *Cell* 2008, **135:**49-60.
117. Zhou L, Sokolskaja E, Jolly C, James W, Cowley SA, Fassati A: **Transportin 3 promotes a nuclear maturation step required for efficient HIV-1 integration.** *PLoS Pathog* 2011, **7:**e1002194.
118. Valle-Casuso JC, Di Nunzio F, Yang Y, Reszka N, Lienlaf M, Arhel N, Perez P, Brass AL, Diaz-Griffero F: **TNPO3 is required for HIV-1 replication after nuclear import but prior to integration and binds the HIV-1 core.** *J Virol* 2012, **86:**5931-5936.
119. Zaitseva L, Cherepanov P, Leyens L, Wilson SJ, Rasaiyaah J, Fassati A: **HIV-1 exploits importin 7 to maximize nuclear import of its DNA genome.** *Retrovirology* 2009, **6:**11.
120. Zielske SP, Stevenson M: **Modest but reproducible inhibition of human immunodeficiency virus type 1 infection in macrophages following LEDGFp75 silencing.** *J Virol* 2006, **80:**7275-7280.

121. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ: **Identification of host proteins required for HIV infection through a functional genomic screen.** *Science* 2008, **319:**921-926.
122. Woodward CL, Prakobwanakit S, Mosessian S, Chow SA: **Integrase interacts with nucleoporin NUP153 to mediate the nuclear import of human immunodeficiency virus type 1.** *J Virol* 2009, **83:**6522-6533.
123. Matreyek KA, Engelman A: **The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection is determined by the viral capsid.** *J Virol* 2011, **85:**7818-7827.
124. Ebina H, Aoki J, Hatta S, Yoshida T, Koyanagi Y: **Role of Nup98 in nuclear entry of human immunodeficiency virus type 1 cDNA.** *Microbes Infect* 2004, **6:**715-724.
125. Engelman A, Mizuuchi K, Craigie R: **HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer.** *Cell* 1991, **67:**1211-1221.
126. Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, Hawkins ME, Brochon JC: **DNA binding induces dissociation of the multimeric form of HIV-1 integrase: a time-resolved fluorescence anisotropy study.** *Proc Natl Acad Sci U S A* 2001, **98:**10090-10095.
127. Faure A, Calmels C, Desjobert C, Castroviejo M, Caumont-Sarcos A, Tarrago-Litvak L, Litvak S, Parissi V: **HIV-1 integrase crosslinked oligomers are active in vitro.** *Nucleic Acids Res* 2005, **33:**977-986.
128. Guiot E, Carayon K, Delelis O, Simon F, Tauc P, Zubin E, Gottikh M, Mouscadet JF, Brochon JC, Deprez E: **Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity.** *J Biol Chem* 2006, **281:**22707-22719.
129. Craigie R: **HIV integrase, a brief overview from chemistry to therapeutics.** *J Biol Chem* 2001, **276:**23213-23216.
130. Bushman FD, Fujiwara T, Craigie R: **Retroviral DNA integration directed by HIV integration protein in vitro.** *Science* 1990, **249:**1555-1558.
131. Li M, Mizuuchi M, Burke TR, Jr., Craigie R: **Retroviral DNA integration: reaction pathway and critical intermediates.** *Embo J* 2006, **25:**1295-1304.
132. Jordan A, Bisgrove D, Verdin E: **HIV reproducibly establishes a latent infection after acute infection of T cells in vitro.** *Embo J* 2003, **22:**1868-1877.
133. Lewinski MK, Bisgrove D, Shinn P, Chen H, Hoffmann C, Hannenhalli S, Verdin E, Berry CC, Ecker JR, Bushman FD: **Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription.** *J Virol* 2005, **79:**6610-6619.
134. Schauer M, Billich A: **The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA-binding.** *Biochem Biophys Res Commun* 1992, **185:**874-880.
135. Vink C, van Gent DC, Plasterk RH: **Integration of human immunodeficiency virus types 1 and 2 DNA in vitro by cytoplasmic extracts of Moloney murine leukemia virus-infected mouse NIH 3T3 cells.** *J Virol* 1990, **64:**5219-5222.
136. Chow SA, Vincent KA, Ellison V, Brown PO: **Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus.** *Science* 1992, **255:**723-726.

137. Brin E, Yi J, Skalka AM, Leis J: **Modeling the late steps in HIV-1 retroviral integrase-catalyzed DNA integration.** *J Biol Chem* 2000, **275:**39287-39295.
138. Yoder KE, Bushman FD: **Repair of gaps in retroviral DNA integration intermediates.** *J Virol* 2000, **74:**11191-11200.
139. Engelman A: **In vivo analysis of retroviral integrase structure and function.** *Adv Virus Res* 1999, **52:**411-426.
140. Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, Kellam P, Cherepanov P, Engelman A: **LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration.** *Genes Dev* 2007, **21:**1767-1778.
141. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR, Bushman FD: **Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences.** *PLoS Biol* 2004, **2:**E234.
142. Carteau S, Batson SC, Poljak L, Mouscadet JF, de Rocquigny H, Darlix JL, Roques BP, Kas E, Auclair C: **Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg2+-dependent DNA integration in vitro.** *J Virol* 1997, **71:**6225-6229.
143. Ciuffi A, Diamond TL, Hwang Y, Marshall HM, Bushman FD: **Modulating target site selection during human immunodeficiency virus DNA integration in vitro with an engineered tethering factor.** *Hum Gene Ther* 2006, **17:**960-967.
144. Gijsbers R, Vets S, De Rijck J, Ocwieja KE, Ronen K, Malani N, Bushman FD, Debyser Z: **Role of the PWWP domain of lens epithelium-derived growth factor (LEDGF)/p75 cofactor in lentiviral integration targeting.** *J Biol Chem* 2011, **286:**41812-41825.
145. Ferris AL, Wu X, Hughes CM, Stewart C, Smith SJ, Milne TA, Wang GG, Shun MC, Allis CD, Engelman A, Hughes SH: **Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration.** *Proc Natl Acad Sci U S A* 2010, **107:**3135-3140.
146. Ocwieja KE, Brady TL, Ronen K, Huegel A, Roth SL, Schaller T, James LC, Towers GJ, Young JA, Chanda SK, et al: **HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2.** *PLoS Pathog* 2011, **7:**e1001313.
147. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, Debyser Z, Engelborghs Y: **LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells.** *J Biol Chem* 2003, **278:**33528-33539.
148. Maertens G, Cherepanov P, Debyser Z, Engelborghs Y, Engelman A: **Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75.** *J Biol Chem* 2004, **279:**33421-33429.
149. Bukovsky A, Gottlinger H: **Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease.** *J Virol* 1996, **70:**6820-6825.
150. Studamire B, Goff SP: **Host proteins interacting with the Moloney murine leukemia virus integrase: multiple transcriptional regulators and chromatin binding factors.** *Retrovirology* 2008, **5:**48.

151. Lee MS, Craigie R: **Protection of retroviral DNA from autointegration: involvement of a cellular factor.** *Proc Natl Acad Sci U S A* 1994, **91:**9823-9827.
152. Farnet CM, Bushman FD: **HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro.** *Cell* 1997, **88:**483-492.
153. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP: **Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5.** *Science* 1994, **266:**2002-2006.
154. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, Tempe D, Vandekerckhove L, Moisant F, Ben-Slama L, Witvrouw M, et al: **Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication.** *J Biol Chem* 2005, **280:**25517-25523.
155. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, De Clercq E, Debyser Z: **HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells.** *J Biol Chem* 2003, **278:**372-381.
156. Lin CW, Engelman A: **The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes.** *J Virol* 2003, **77:**5030-5036.
157. Raghavendra NK, Shkriabai N, Graham R, Hess S, Kvaratskhelia M, Wu L: **Identification of host proteins associated with HIV-1 preintegration complexes isolated from infected CD4+ cells.** *Retrovirology* 2010, **7:**66.
158. Chen H, Engelman A: **The barrier-to-autointegration protein is a host factor for HIV type 1 integration.** *Proc Natl Acad Sci U S A* 1998, **95:**15270-15274.
159. Singh DP, Ohguro N, Kikuchi T, Sueno T, Reddy VN, Yuge K, Chylack LT, Jr., Shinohara T: **Lens epithelium-derived growth factor: effects on growth and survival of lens epithelial cells, keratinocytes, and fibroblasts.** *Biochem Biophys Res Commun* 2000, **267:**373-381.
160. Singh DP, Ohguro N, Chylack LT, Jr., Shinohara T: **Lens epithelium-derived growth factor: increased resistance to thermal and oxidative stresses.** *Invest Ophthalmol Vis Sci* 1999, **40:**1444-1451.
161. Ganapathy V, Daniels T, Casiano CA: **LEDGF/p75: a novel nuclear autoantigen at the crossroads of cell survival and apoptosis.** *Autoimmun Rev* 2003, **2:**290-297.
162. Izumoto Y, Kuroda T, Harada H, Kishimoto T, Nakamura H: **Hepatoma-derived growth factor belongs to a gene family in mice showing significant homology in the amino terminus.** *Biochem Biophys Res Commun* 1997, **238:**26-32.
163. Ikegame K, Yamamoto M, Kishima Y, Enomoto H, Yoshida K, Suemura M, Kishimoto T, Nakamura H: **A new member of a hepatoma-derived growth factor gene family can translocate to the nucleus.** *Biochem Biophys Res Commun* 1999, **266:**81-87.
164. Dietz F, Franken S, Yoshida K, Nakamura H, Kappler J, Gieselmann V: **The family of hepatoma-derived growth factor proteins: characterization of a new member HRP-4 and classification of its subfamilies.** *Biochem J* 2002, **366:**491-500.

165. Vanegas M, Llano M, Delgado S, Thompson D, Peretz M, Poeschla E: **Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering.** *J Cell Sci* 2005, **118:**1733-1743.
166. Gijsbers R, Ronen K, Vets S, Malani N, De Rijck J, McNeely M, Bushman FD, Debyser Z: **LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin.** *Mol Ther* 2010, **18:**552-560.
167. Vandegraaff N, Devroe E, Turlure F, Silver PA, Engelman A: **Biochemical and genetic analyses of integrase-interacting proteins lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication.** *Virology* 2006, **346:**415-426.
168. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH, Teo W, Poeschla EM: **An essential role for LEDGF/p75 in HIV integration.** *Science* 2006, **314:**461-464.
169. Cherepanov P, Sun ZY, Rahman S, Maertens G, Wagner G, Engelman A: **Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75.** *Nat Struct Mol Biol* 2005, **12:**526-532.
170. Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, Peretz M, Poeschla EM: **LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes.** *J Virol* 2004, **78:**9524-9537.
171. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A: **Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75.** *Proc Natl Acad Sci U S A* 2005, **102:**17308-17313.
172. Rahman S, Lu R, Vandegraaff N, Cherepanov P, Engelman A: **Structure-based mutagenesis of the integrase-LEDGF/p75 interface uncouples a strict correlation between in vitro protein binding and HIV-1 fitness.** *Virology* 2007, **357:**79-90.
173. Busschots K, Voet A, De Maeyer M, Rain JC, Emiliani S, Benarous R, Desender L, Debyser Z, Christ F: **Identification of the LEDGF/p75 binding site in HIV-1 integrase.** *J Mol Biol* 2007, **365:**1480-1492.
174. Pandey KK, Sinha S, Grandgenett DP: **Transcriptional coactivator LEDGF/p75 modulates human immunodeficiency virus type 1 integrase-mediated concerted integration.** *J Virol* 2007, **81:**3969-3979.
175. Llano M, Delgado S, Vanegas M, Poeschla EM: **Lens epithelium-derived growth factor/p75 prevents proteasomal degradation of HIV-1 integrase.** *J Biol Chem* 2004, **279:**55570-55577.
176. Engelman A, Cherepanov P: **The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication.** *PLoS Pathog* 2008, **4:**e1000046.
177. Schrijvers R, De Rijck J, Demeulemeester J, Adachi N, Vets S, Ronen K, Christ F, Bushman FD, Debyser Z, Gijsbers R: **LEDGF/p75-independent HIV-1 replication demonstrates a role for HRP-2 and remains sensitive to inhibition by LEDGINs.** *PLoS Pathog* 2012, **8:**e1002558.
178. De Rijck J, Vandekerckhove L, Gijsbers R, Hombrouck A, Hendrix J, Vercammen J, Engelborghs Y, Christ F, Debyser Z: **Overexpression of the lens**

epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. *J Virol* 2006, **80:**11498-11509.
179. Aravind L, Koonin EV: **Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system.** *Genome Res* 2001, **11:**1365-1374.
180. Downs JA, Jackson SP: **A means to a DNA end: the many roles of Ku.** *Nat Rev Mol Cell Biol* 2004, **5:**367-378.
181. Doherty AJ, Jackson SP, Weller GR: **Identification of bacterial homologues of the Ku DNA repair proteins.** *FEBS Lett* 2001, **500:**186-188.
182. Wu X, Lieber MR: **Protein-protein and protein-DNA interaction regions within the DNA end-binding protein Ku70-Ku86.** *Mol Cell Biol* 1996, **16:**5186-5193.
183. Koike M, Miyasaka T, Mimori T, Shiomi T: **Subcellular localization and protein-protein interaction regions of Ku proteins.** *Biochem Biophys Res Commun* 1998, **252:**679-685.
184. Koike M, Ikuta T, Miyasaka T, Shiomi T: **The nuclear localization signal of the human Ku70 is a variant bipartite type recognized by the two components of nuclear pore-targeting complex.** *Exp Cell Res* 1999, **250:**401-413.
185. Chan YG, Cardwell MM, Hermanas TM, Uchiyama T, Martinez JJ: **Rickettsial outer-membrane protein B (rOmpB) mediates bacterial invasion through Ku70 in an actin, c-Cbl, clathrin and caveolin 2-dependent manner.** *Cell Microbiol* 2009, **11:**629-644.
186. Prabhakar BS, Allaway GP, Srinivasappa J, Notkins AL: **Cell surface expression of the 70-kD component of Ku, a DNA-binding nuclear autoantigen.** *J Clin Invest* 1990, **86:**1301-1305.
187. Dalziel RG, Mendelson SC, Quinn JP: **The nuclear autoimmune antigen Ku is also present on the cell surface.** *Autoimmunity* 1992, **13:**265-267.
188. Zhang X, Brann TW, Zhou M, Yang J, Oguariri RM, Lidie KB, Imamichi H, Huang DW, Lempicki RA, Baseler MW, et al: **Cutting edge: Ku70 is a novel cytosolic DNA sensor that induces type III rather than type I IFN.** *J Immunol* 2011, **186:**4541-4545.
189. Subramanian C, Opipari AW, Jr., Bian X, Castle VP, Kwok RP: **Ku70 acetylation mediates neuroblastoma cell death induced by histone deacetylase inhibitors.** *Proc Natl Acad Sci U S A* 2005, **102:**4842-4847.
190. Bakalkin G, Yakovleva T, Hurd YL, Nussenzweig A, Li GC, Terenius L: **Autoantigen Ku in the brain. Developmentally regulated expression and subcellular localization.** *Neuroreport* 1998, **9:**2147-2151.
191. Dvir A, Peterson SR, Knuth MW, Lu H, Dynan WS: **Ku autoantigen is the regulatory component of a template-associated protein kinase that phosphorylates RNA polymerase II.** *Proc Natl Acad Sci U S A* 1992, **89:**11920-11924.
192. Cohen HY, Lavu S, Bitterman KJ, Hekking B, Imahiyerobo TA, Miller C, Frye R, Ploegh H, Kessler BM, Sinclair DA: **Acetylation of the C terminus of Ku70 by CBP and PCAF controls Bax-mediated apoptosis.** *Mol Cell* 2004, **13:**627-638.

193. Subramanian C, Jarzembowski JA, Opipari AW, Jr., Castle VP, Kwok RP: **HDAC6 deacetylates Ku70 and regulates Ku70-Bax binding in neuroblastoma.** *Neoplasia* 2011, **13:**726-734.
194. Lees-Miller SP, Meek K: **Repair of DNA double strand breaks by non-homologous end joining.** *Biochimie* 2003, **85:**1161-1173.
195. Tuteja R, Tuteja N: **Ku autoantigen: a multifunctional DNA-binding protein.** *Crit Rev Biochem Mol Biol* 2000, **35:**1-33.
196. Giffin W, Torrance H, Rodda DJ, Prefontaine GG, Pope L, Hache RJ: **Sequence-specific DNA binding by Ku autoantigen and its effects on transcription.** *Nature* 1996, **380:**265-268.
197. Walker JR, Corpina RA, Goldberg J: **Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair.** *Nature* 2001, **412:**607-614.
198. Fisher TS, Zakian VA: **Ku: a multifunctional protein involved in telomere maintenance.** *DNA Repair (Amst)* 2005, **4:**1215-1226.
199. Chou CH, Wang J, Knuth MW, Reeves WH: **Role of a major autoepitope in forming the DNA binding site of the p70 (Ku) antigen.** *J Exp Med* 1992, **175:**1677-1684.
200. Aravind L, Koonin EV: **SAP - a putative DNA-binding motif involved in chromosomal organization.** *Trends Biochem Sci* 2000, **25:**112-114.
201. Wang J, Dong X, Myung K, Hendrickson EA, Reeves WH: **Identification of two domains of the p70 Ku protein mediating dimerization with p80 and DNA binding.** *J Biol Chem* 1998, **273:**842-848.
202. Roe T, Chow SA, Brown PO: **3'-end processing and kinetics of 5'-end joining during retroviral integration in vivo.** *J Virol* 1997, **71:**1334-1340.
203. Lau A, Swinbank KM, Ahmed PS, Taylor DL, Jackson SP, Smith GC, O'Connor MJ: **Suppression of HIV-1 infection by a small molecule inhibitor of the ATM kinase.** *Nat Cell Biol* 2005, **7:**493-500.
204. Daniel R, Pomerantz RJ: **ATM: HIV-1's Achilles heel?** *Nat Cell Biol* 2005, **7:**452-453.
205. Skalka AM, Katz RA: **Retroviral DNA integration and the DNA damage response.** *Cell Death Differ* 2005, **12 Suppl 1:**971-978.
206. Daniel R, Greger JG, Katz RA, Taganov KD, Wu X, Kappes JC, Skalka AM: **Evidence that stable retroviral transduction and cell survival following DNA integration depend on components of the nonhomologous end joining repair pathway.** *J Virol* 2004, **78:**8573-8581.
207. Daniel R, Katz RA, Merkel G, Hittle JC, Yen TJ, Skalka AM: **Wortmannin potentiates integrase-mediated killing of lymphocytes and reduces the efficiency of stable transduction by retroviruses.** *Mol Cell Biol* 2001, **21:**1164-1172.
208. Jeanson L, Subra F, Vaganay S, Hervy M, Marangoni E, Bourhis J, Mouscadet JF: **Effect of Ku80 depletion on the preintegrative steps of HIV-1 replication in human cells.** *Virology* 2002, **300:**100-108.
209. Daniel R, Katz RA, Skalka AM: **A role for DNA-PK in retroviral DNA integration.** *Science* 1999, **284:**644-647.

210. Greene WC, Peterlin BM: **Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy.** *Nat Med* 2002, **8:**673-680.
211. Li L, Olvera JM, Yoder KE, Mitchell RS, Butler SL, Lieber M, Martin SL, Bushman FD: **Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection.** *Embo J* 2001, **20:**3272-3281.
212. Jeanson L, Mouscadet JF: **Ku represses the HIV-1 transcription: identification of a putative Ku binding site homologous to the mouse mammary tumor virus NRE1 sequence in the HIV-1 long terminal repeat.** *J Biol Chem* 2002, **277:**4918-4924.
213. Manganaro L, Lusic M, Gutierrez MI, Cereseto A, Del Sal G, Giacca M: **Concerted action of cellular JNK and Pin1 restricts HIV-1 genome integration to activated CD4+ T lymphocytes.** *Nat Med* 2010, **16:**329-333.
214. Zamborlini A, Coiffic A, Beauclair G, Delelis O, Paris J, Koh Y, Magne F, Giron ML, Tobaly-Tapiero J, Deprez E, et al: **Impairment of human immunodeficiency virus type-1 integrase SUMOylation correlates with an early replication defect.** *J Biol Chem* 2011, **286:**21013-21022.
215. Zheng Y, Ao Z, Wang B, Jayappa KD, Yao X: **Host protein Ku70 binds and protects HIV-1 integrase from proteasomal degradation and is required for HIV replication.** *J Biol Chem* 2011, **286:**17722-17735.
216. Mulder LC, Muesing MA: **Degradation of HIV-1 integrase by the N-end rule pathway.** *J Biol Chem* 2000, **275:**29749-29753.
217. Baboshina OV, Haas AL: **Novel multiubiquitin chain linkages catalyzed by the conjugating enzymes E2EPF and RAD6 are recognized by 26 S proteasome subunit 5.** *J Biol Chem* 1996, **271:**2823-2831.
218. Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH: **The origins of acquired immune deficiency syndrome viruses: where and when?** *Philos Trans R Soc Lond B Biol Sci* 2001, **356:**867-876.
219. Spence J, Sadis S, Haas AL, Finley D: **A ubiquitin mutant with specific defects in DNA repair and multiubiquitination.** *Mol Cell Biol* 1995, **15:**1265-1273.
220. Tasaki T, Mulder LC, Iwamatsu A, Lee MJ, Davydov IV, Varshavsky A, Muesing M, Kwon YT: **A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons.** *Mol Cell Biol* 2005, **25:**7120-7136.
221. Ravid T, Hochstrasser M: **Diversity of degradation signals in the ubiquitin-proteasome system.** *Nat Rev Mol Cell Biol* 2008, **9:**679-690.
222. Sriram SM, Kim BY, Kwon YT: **The N-end rule pathway: emerging functions and molecular principles of substrate recognition.** *Nat Rev Mol Cell Biol* 2011, **12:**735-747.
223. Kaelin WG, Jr.: **The von Hippel-Lindau tumour suppressor protein: O2 sensing and cancer.** *Nat Rev Cancer* 2008, **8:**865-873.
224. Lisztwan J, Imbert G, Wirbelauer C, Gstaiger M, Krek W: **The von Hippel-Lindau tumor suppressor protein is a component of an E3 ubiquitin-protein ligase activity.** *Genes Dev* 1999, **13:**1822-1833.
225. Debigare R, Price SR: **Proteolysis, the ubiquitin-proteasome system, and renal diseases.** *Am J Physiol Renal Physiol* 2003, **285:**F1-8.

226. Mulder LC, Chakrabarti LA, Muesing MA: **Interaction of HIV-1 integrase with DNA repair protein hRad18.** *J Biol Chem* 2002, **277:**27489-27493.
227. Tateishi S, Sakuraba Y, Masuyama S, Inoue H, Yamaizumi M: **Dysfunction of human Rad18 results in defective postreplication repair and hypersensitivity to multiple mutagens.** *Proc Natl Acad Sci U S A* 2000, **97:**7927-7932.
228. Putnam CD, Hayes TK, Kolodner RD: **Post-replication repair suppresses duplication-mediated genome instability.** *PLoS Genet* 2010, **6:**e1000933.
229. Geng L, Huntoon CJ, Karnitz LM: **RAD18-mediated ubiquitination of PCNA activates the Fanconi anemia DNA repair network.** *J Cell Biol* 2010, **191:**249-257.
230. Yang K, Moldovan GL, D'Andrea AD: **RAD18-dependent recruitment of SNM1A to DNA repair complexes by a ubiquitin-binding zinc finger.** *J Biol Chem* 2010, **285:**19085-19091.
231. Schwartz DC, Hochstrasser M: **A superfamily of protein tags: ubiquitin, SUMO and related modifiers.** *Trends Biochem Sci* 2003, **28:**321-328.
232. Matic I, Schimmel J, Hendriks IA, van Santen MA, van de Rijke F, van Dam H, Gnad F, Mann M, Vertegaal AC: **Site-specific identification of SUMO-2 targets in cells reveals an inverted SUMOylation motif and a hydrophobic cluster SUMOylation motif.** *Mol Cell* 2010, **39:**641-652.
233. Seeler JS, Dejean A: **Nuclear and unclear functions of SUMO.** *Nat Rev Mol Cell Biol* 2003, **4:**690-699.
234. Geiss-Friedlander R, Melchior F: **Concepts in sumoylation: a decade on.** *Nat Rev Mol Cell Biol* 2007, **8:**947-956.
235. Guo D, Li M, Zhang Y, Yang P, Eckenrode S, Hopkins D, Zheng W, Purohit S, Podolsky RH, Muir A, et al: **A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes.** *Nat Genet* 2004, **36:**837-841.
236. Saitoh H, Hinchey J: **Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3.** *J Biol Chem* 2000, **275:**6252-6258.
237. Arriagada G, Muntean LN, Goff SP: **SUMO-interacting motifs of human TRIM5alpha are important for antiviral activity.** *PLoS Pathog* 2011, **7:**e1002019.
238. Gareau JR, Reverter D, Lima CD: **Determinants of small ubiquitin-like modifier 1 (SUMO1) protein specificity, E3 ligase, and SUMO-RanGAP1 binding activities of nucleoporin RanBP2.** *J Biol Chem* 2012, **287:**4740-4751.
239. Hecker CM, Rabiller M, Haglund K, Bayer P, Dikic I: **Specification of SUMO1- and SUMO2-interacting motifs.** *J Biol Chem* 2006, **281:**16117-16127.
240. Song J, Durrin LK, Wilkinson TA, Krontiris TG, Chen Y: **Identification of a SUMO-binding motif that recognizes SUMO-modified proteins.** *Proc Natl Acad Sci U S A* 2004, **101:**14373-14378.
241. Song J, Zhang Z, Hu W, Chen Y: **Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation.** *J Biol Chem* 2005, **280:**40122-40129.
242. Kerscher O: **SUMO junction-what's your function? New insights through SUMO-interacting motifs.** *EMBO Rep* 2007, **8:**550-555.

243. Pichler A, Knipscheer P, Oberhofer E, van Dijk WJ, Korner R, Olsen JV, Jentsch S, Melchior F, Sixma TK: **SUMO modification of the ubiquitin-conjugating enzyme E2-25K.** *Nat Struct Mol Biol* 2005, **12:**264-269.
244. Hoege C, Pfander B, Moldovan GL, Pyrowolakis G, Jentsch S: **RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO.** *Nature* 2002, **419:**135-141.
245. Zhang R, Mehla R, Chauhan A: **Perturbation of host nuclear membrane component RanBP2 impairs the nuclear import of human immunodeficiency virus -1 preintegration complex (DNA).** *PLoS One* 2010, **5:**e15620.
246. Yurchenko V, Xue Z, Gama V, Matsuyama S, Sadofsky MJ: **Ku70 is stabilized by increased cellular SUMO.** *Biochem Biophys Res Commun* 2008, **366:**263-268.
247. Bueno MT, Garcia-Rivera JA, Kugelman JR, Morales E, Rosas-Acosta G, Llano M: **SUMOylation of the lens epithelium-derived growth factor/p75 attenuates its transcriptional activity on the heat shock protein 27 promoter.** *J Mol Biol* 2010, **399:**221-239.
248. Garcia-Wilson E, Perkins ND: **p21WAF1/CIP1 regulates the p300 sumoylation motif CRD1 through a C-terminal domain independently of cyclin/CDK binding.** *Cell Cycle* 2005, **4:**1113-1119.
249. Altmannova V, Eckert-Boulet N, Arneric M, Kolesar P, Chaloupkova R, Damborsky J, Sung P, Zhao X, Lisby M, Krejci L: **Rad52 SUMOylation affects the efficiency of the DNA repair.** *Nucleic Acids Res* 2010, **38:**4708-4721.
250. Terreni M, Valentini P, Liverani V, Gutierrez MI, Di Primio C, Di Fenza A, Tozzini V, Allouch A, Albanese A, Giacca M, Cereseto A: **GCN5-dependent acetylation of HIV-1 integrase enhances viral integration.** *Retrovirology* 2010, **7:**18.
251. Cereseto A, Manganaro L, Gutierrez MI, Terreni M, Fittipaldi A, Lusic M, Marcello A, Giacca M: **Acetylation of HIV-1 integrase by p300 regulates viral integration.** *Embo J* 2005, **24:**3070-3081.
252. Legube G, Trouche D: **Regulating histone acetyltransferases and deacetylases.** *EMBO Rep* 2003, **4:**944-947.
253. Topper M, Luo Y, Zhadina M, Mohammed K, Smith L, Muesing MA: **Posttranslational acetylation of the human immunodeficiency virus type 1 integrase carboxyl-terminal domain is dispensable for viral replication.** *J Virol* 2007, **81:**3012-3017.
254. Ubersax JA, Ferrell JE, Jr.: **Mechanisms of specificity in protein phosphorylation.** *Nat Rev Mol Cell Biol* 2007, **8:**530-541.
255. De Luca L, De Grazia S, Ferro S, Gitto R, Christ F, Debyser Z, Chimirri A: **HIV-1 integrase strand-transfer inhibitors: design, synthesis and molecular modeling investigation.** *Eur J Med Chem* 2011, **46:**756-764.
256. Maltez F, Doroana M, Branco T, Valente C: **Recent advances in antiretroviral treatment and prevention in HIV-infected patients.** *Curr Opin HIV AIDS* 2011, **6 Suppl 1:**S21-30.
257. Flexner C: **HIV drug development: the next 25 years.** *Nat Rev Drug Discov* 2007, **6:**959-966.

258. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA: **Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor.** *Science* 1992, **256:**1783-1790.
259. Wlodawer A, Erickson JW: **Structure-based inhibitors of HIV-1 protease.** *Annu Rev Biochem* 1993, **62:**543-585.
260. Gupta K, Diamond T, Hwang Y, Bushman F, Van Duyne GD: **Structural properties of HIV integrase. Lens epithelium-derived growth factor oligomers.** *J Biol Chem* 2010, **285:**20303-20315.
261. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR: **Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases.** *Science* 1994, **266:**1981-1986.
262. McColl DJ, Chen X: **Strand transfer inhibitors of HIV-1 integrase: bringing IN a new era of antiretroviral therapy.** *Antiviral Res* 2010, **85:**101-118.
263. Blanco JL, Varghese V, Rhee SY, Gatell JM, Shafer RW: **HIV-1 integrase inhibitor resistance and its clinical implications.** *J Infect Dis* 2011, **203:**1204-1214.
264. Zinzalla G, Thurston DE: **Targeting protein-protein interactions for therapeutic intervention: a challenge for the future.** *Future Med Chem* 2009, **1:**65-93.
265. Zhang ZY, Poorman RA, Maggiora LL, Heinrikson RL, Kezdy FJ: **Dissociative inhibition of dimeric enzymes. Kinetic characterization of the inhibition of HIV-1 protease by its COOH-terminal tetrapeptide.** *J Biol Chem* 1991, **266:**15591-15594.
266. Zutshi R, Chmielewski J: **Targeting the dimerization interface for irreversible inhibition of HIV-1 protease.** *Bioorg Med Chem Lett* 2000, **10:**1901-1903.
267. Shultz MD, Ham YW, Lee SG, Davis DA, Brown C, Chmielewski J: **Small-molecule dimerization inhibitors of wild-type and mutant HIV protease: a focused library approach.** *J Am Chem Soc* 2004, **126:**9886-9887.
268. McNeely M, Hendrix J, Busschots K, Boons E, Deleersnijder A, Gerard M, Christ F, Debyser Z: **In vitro DNA tethering of HIV-1 integrase by the transcriptional coactivator LEDGF/p75.** *J Mol Biol* 2011, **410:**811-830.
269. Michel F, Crucifix C, Granger F, Eiler S, Mouscadet JF, Korolev S, Agapkina J, Ziganshin R, Gottikh M, Nazabal A, et al: **Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor.** *Embo J* 2009, **28:**980-991.
270. Vets S, Kimpel J, Volk A, De Rijck J, Schrijvers R, Verbinnen B, Maes W, Von Laer D, Debyser Z, Gijsbers R: **Lens epithelium-derived growth factor/p75 qualifies as a target for HIV gene therapy in the NSG mouse model.** *Mol Ther* 2012, **20:**908-917.
271. Hayouka Z, Rosenbluh J, Levin A, Loya S, Lebendiker M, Veprintsev D, Kotler M, Hizi A, Loyter A, Friedler A: **Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium.** *Proc Natl Acad Sci U S A* 2007, **104:**8316-8321.
272. Hayouka Z, Levin A, Maes M, Hadas E, Shalev DE, Volsky DJ, Loyter A, Friedler A: **Mechanism of action of the HIV-1 integrase inhibitory peptide LEDGF 361-370.** *Biochem Biophys Res Commun* 2010, **394:**260-265.

273. De Luca L, Gitto R, Christ F, Ferro S, De Grazia S, Morreale F, Debyser Z, Chimirri A: **4-[1-(4-Fluorobenzyl)-4-hydroxy-1H-indol-3-yl]-2-hydroxy-4-oxobut-2-enoic acid as a prototype to develop dual inhibitors of HIV-1 integration process.** *Antiviral Res* 2011, **92:**102-107.
274. Christ F, Voet A, Marchand A, Nicolet S, Desimmie BA, Marchand D, Bardiot D, Van der Veken NJ, Van Remoortel B, Strelkov SV, et al: **Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication.** *Nat Chem Biol* 2010, **6:**442-448.
275. Yung E, Sorin M, Pal A, Craig E, Morozov A, Delattre O, Kappes J, Ott D, Kalpana GV: **Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1.** *Nat Med* 2001, **7:**920-926.
276. Zhao L, O'Reilly MK, Shultz MD, Chmielewski J: **Interfacial peptide inhibitors of HIV-1 integrase activity and dimerization.** *Bioorg Med Chem Lett* 2003, **13:**1175-1177.
277. Maroun RG, Gayet S, Benleulmi MS, Porumb H, Zargarian L, Merad H, Leh H, Mouscadet JF, Troalen F, Fermandjian S: **Peptide inhibitors of HIV-1 integrase dissociate the enzyme oligomers.** *Biochemistry* 2001, **40:**13840-13848.
278. Yao XJ, Subbramanian RA, Rougeau N, Boisvert F, Bergeron D, Cohen EA: **Mutagenic analysis of human immunodeficiency virus type 1 Vpr: role of a predicted N-terminal alpha-helical structure in Vpr nuclear localization and virion incorporation.** *J Virol* 1995, **69:**7032-7044.
279. Yao X-J, Lemay J, Rougeau N, Clément M, Kurtz S, Belhumeur P, Cohen EA: **Genetic Selection of peptide inhibitors of human immunodeficiency virus type 1 (HIV-1) Vpr.** *J Biol Chem* 2002, **277:**48816-48826.
280. Yao XJ, Rougeau N, Duisit G, Lemay J, Cohen EA: **Analysis of HIV-1 Vpr determinants responsible for cell growth arrest in Saccharomyces cerevisiae.** *Retrovirology* 2004, **1:**21.
281. Gietz D, St Jean A, Woods RA, Schiestl RH: **Improved method for high efficiency transformation of intact yeast cells.** *Nucleic Acids Res* 1992, **20:**1425.
282. Ao Z, Fowke KR, Cohen EA, Yao X: **Contribution of the C-terminal tri-lysine regions of human immunodeficiency virus type 1 integrase for efficient reverse transcription and viral DNA nuclear import.** *Retrovirology* 2005, **2:**62.
283. Ao Z, Danappa Jayappa K, Wang B, Zheng Y, Kung S, Rassart E, Depping R, Kohler M, Cohen EA, Yao X: **Importin alpha3 interacts with HIV-1 integrase and contributes to HIV-1 nuclear import and replication.** *J Virol* 2010, **84:**8650-8663.
284. Yao XJ, Mouland AJ, Subbramanian RA, Forget J, Rougeau N, Bergeron D, Cohen EA: **Vpr stimulates viral expression and induces cell killing in human immunodeficiency virus type 1-infected dividing Jurkat T cells.** *J Virol* 1998, **72:**4686-4693.
285. Zheng Y, Ao Z, Jayappa KD, Yao X: **Characterization of the HIV-1 integrase chromatin- and LEDGF/p75-binding abilities by mutagenic analysis within the catalytic core domain of integrase.** *Virol J* 2010, **7:**68.
286. Sorin M, Yung E, Wu X, Kalpana GV: **HIV-1 replication in cell lines harboring INI1/hSNF5 mutations.** *Retrovirology* 2006, **3:**56.

287. Strack B, Calistri A, Gottlinger HG: **Late assembly domain function can exhibit context dependence and involves ubiquitin residues implicated in endocytosis.** *J Virol* 2002, **76:**5472-5479.
288. Kamitani T, Nguyen HP, Kito K, Fukuda-Kamitani T, Yeh ET: **Covalent modification of PML by the sentrin family of ubiquitin-like proteins.** *J Biol Chem* 1998, **273:**3117-3120.
289. Liang C, Stillman B: **Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants.** *Genes Dev* 1997, **11:**3375-3386.
290. Llano M, Vanegas M, Hutchins N, Thompson D, Delgado S, Poeschla EM: **Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75.** *J Mol Biol* 2006, **360:**760-773.
291. Sarge KD, Park-Sarge OK: **Detection of proteins sumoylated in vivo and in vitro.** *Methods Mol Biol* 2009, **590:**265-277.
292. Xu Z, Zheng Y, Ao Z, Clement M, Mouland AJ, Kalpana GV, Belhumeur P, Cohen EA, Yao X: **Contribution of the C-terminal region within the catalytic core domain of HIV-1 integrase to yeast lethality, chromatin binding and viral replication.** *Retrovirology* 2008, **5:**102.
293. Mouland AJ, Mercier J, Luo M, Bernier L, DesGroseillers L, Cohen EA: **The double-stranded RNA-binding protein Staufen is incorporated in human immunodeficiency virus type 1: evidence for a role in genomic RNA encapsidation.** *J Virol* 2000, **74:**5441-5451.
294. Malnati MS, Scarlatti G, Gatto F, Salvatori F, Cassina G, Rutigliano T, Volpi R, Lusso P: **A universal real-time PCR assay for the quantification of group-M HIV-1 proviral load.** *Nat Protoc* 2008, **3:**1240-1248.
295. Parissi V, Caumont AB, de Soultrait VR, Calmels C, Pichuantes S, Litvak S, Dupont CH: **Selection of amino acid substitutions restoring activity of HIV-1 integrase mutated in its catalytic site using the yeast Saccharomyces cerevisiae.** *J Mol Biol* 2000, **295:**755-765.
296. Caumont AB, Jamieson GA, Pichuantes S, Nguyen AT, Litvak S, Dupont C: **Expression of functional HIV-1 integrase in the yeast Saccharomyces cerevisiae leads to the emergence of a lethal phenotype: potential use for inhibitor screening.** *Curr Genet* 1996, **29:**503-510.
297. Parissi V, Caumont A, de Soultrait VR, Desjobert C, Calmels C, Fournier M, Gourgue G, Bonneu M, Tarrago-Litvak L, Litvak S: **The lethal phenotype observed after HIV-1 integrase expression in yeast cells is related to DNA repair and recombination events.** *Gene* 2003, **322:**157-168.
298. Calmels C, de Soultrait VR, Caumont A, Desjobert C, Faure A, Fournier M, Tarrago-Litvak L, Parissi V: **Biochemical and random mutagenesis analysis of the region carrying the catalytic E152 amino acid of HIV-1 integrase.** *Nucleic Acids Res* 2004, **32:**1527-1538.
299. Desfarges S SFJ, Fournier M, Calmels C, Caumont-Sarcos A, Litvak S, Sung P, Parissi V: **Chromosomal integration of LTR-flanked DNA in yeast expressing HIV-1 integrase: down regulation by RAD51. .** *Nucleic Acids Res* 2006, **34:**6215-6224.

300. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, Ecker JR, Bushman F: **A role for LEDGF/p75 in targeting HIV DNA integration.** *Nat Med* 2005, **11:**1287-1289.
301. Parissi V, Caumont A, Richard de Soultrait V, Dupont CH, Pichuantes S, Litvak S: **Inactivation of the SNF5 transcription factor gene abolishes the lethal phenotype induced by the expression of HIV-1 integrase in yeast.** *Gene* 2000, **247:**129-136.
302. Wang W, Cote J, Xue Y, Zhou S, Khavari PA, Biggar SR, Muchardt C, Kalpana GV, Goff SP, Yaniv M, et al: **Purification and biochemical heterogeneity of the mammalian SWI-SNF complex.** *Embo J* 1996, **15:**5370-5382.
303. Ariumi Y, Serhan F, Turelli P, Telenti A, Trono D: **The integrase interactor 1 (INI1) proteins facilitate Tat-mediated human immunodeficiency virus type 1 transcription.** *Retrovirology* 2006, **3:**47.
304. Cannon PM, Wilson W, Byles E, Kingsman SM, Kingsman AJ: **Human immunodeficiency virus type 1 integrase: effect on viral replication of mutations at highly conserved residues.** *J Virol* 1994, **68:**4768-4775.
305. Limon A, Devroe E, Lu R, Ghory HZ, Silver PA, Engelman A: **Nuclear localization of human immunodeficiency virus type 1 preintegration complexes (PICs): V165A and R166A are pleiotropic integrase mutants primarily defective for integration, not PIC nuclear import.** *J Virol* 2002, **76:**10598-10607.
306. Wiskerchen M, Muesing MA: **Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells.** *J Virol* 1995, **69:**376-386.
307. Feller A, Ramos F, Pierard A, Dubois E: **In Saccharomyces cerevisae, feedback inhibition of homocitrate synthase isoenzymes by lysine modulates the activation of LYS gene expression by Lys14p.** *Eur J Biochem* 1999, **261:**163-170.
308. Ikeda T, Nishitsuji H, Zhou X, Nara N, Ohashi T, Kannagi M, Masuda T: **Evaluation of the functional involvement of human immunodeficiency virus type 1 integrase in nuclear import of viral cDNA during acute infection.** *J Virol* 2004, **78:**11563-11573.
309. McKee CJ, Kessl JJ, Shkriabai N, Dar MJ, Engelman A, Kvaratskhelia M: **Dynamic modulation of HIV-1 integrase structure and function by cellular LEDGF protein.** *J Biol Chem* 2008.
310. Parissi V, Calmels C, De Soultrait VR, Caumont A, Fournier M, Chaignepain S, Litvak S: **Functional interactions of human immunodeficiency virus type 1 integrase with human and yeast HSP60.** *J Virol* 2001, **75:**11344-11353.
311. Dolan J, Chen A, Weber IT, Harrison RW, Leis J: **Defining the DNA substrate binding sites on HIV-1 integrase.** *J Mol Biol* 2009, **385:**568-579.
312. Merad H, Porumb H, Zargarian L, Rene B, Hobaika Z, Maroun RG, Mauffret O, Fermandjian S: **An unusual helix turn helix motif in the catalytic core of HIV-1 integrase binds viral DNA and LEDGF.** *PLoS ONE* 2009, **4:**e4081.

313. Engelman A, Hickman AB, Craigie R: **The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding.** *J Virol* 1994, **68:**5911-5917.
314. Heuer TS, Brown PO: **Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex.** *Biochemistry* 1998, **37:**6667-6678.
315. Podtelezhnikov AA, Gao K, Bushman FD, McCammon JA: **Modeling HIV-1 integrase complexes based on their hydrodynamic properties.** *Biopolymers* 2003, **68:**110-120.
316. Drake RR, Neamati N, Hong H, Pilon AA, Sunthankar P, Hume SD, Milne GW, Pommier Y: **Identification of a nucleotide binding site in HIV-1 integrase.** *Proc Natl Acad Sci U S A* 1998, **95:**4170-4175.
317. Jenkins TM, Esposito D, Engelman A, Craigie R: **Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking.** *Embo J* 1997, **16:**6849-6859.
318. Savarino A: **In-Silico docking of HIV-1 integrase inhibitors reveals a novel drug type acting on an enzyme/DNA reaction intermediate.** *Retrovirology* 2007, **4:**21.
319. Shun MC, Botbol Y, Li X, Di Nunzio F, Daigle JE, Yan N, Lieberman J, Lavigne M, Engelman A: **Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity.** *J Virol* 2008, **82:**11555-11567.
320. Al-Mawsawi LQ, Hombrouck A, Dayam R, Debyser Z, Neamati N: **Four-tiered pi interaction at the dimeric interface of HIV-1 integrase critical for DNA integration and viral infectivity.** *Virology* 2008, **377:**355-363.
321. Priet S, Navarro JM, Querat G, Sire J: **Reversion of the lethal phenotype of an HIV-1 integrase mutant virus by overexpression of the same integrase mutant protein.** *J Biol Chem* 2003, **278:**20724-20730.
322. Sayasith K, Sauve G, Yelle J: **Characterization of mutant HIV-1 integrase carrying amino acid changes in the catalytic domain.** *Mol Cells* 2000, **10:**525-532.
323. Amsel AD, Rathaus M, Kronman N, Cohen HY: **Regulation of the proapoptotic factor Bax by Ku70-dependent deubiquitylation.** *Proc Natl Acad Sci U S A* 2008, **105:**5117-5122.
324. Jin S, Weaver DT: **Double-strand break repair by Ku70 requires heterodimerization with Ku80 and DNA binding functions.** *Embo J* 1997, **16:**6874-6885.
325. Koike M, Shiomi T, Koike A: **Dimerization and nuclear localization of ku proteins.** *J Biol Chem* 2001, **276:**11167-11173.
326. Kiel C, Serrano L: **The ubiquitin domain superfold: structure-based sequence alignments and characterization of binding epitopes.** *J Mol Biol* 2006, **355:**821-844.
327. Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, Roelofs J, Finley D, Gygi SP: **A proteomics approach to understanding protein ubiquitination.** *Nat Biotechnol* 2003, **21:**921-926.

328. Dubiel W, Gordon C: **Ubiquitin pathway: another link in the polyubiquitin chain?** *Curr Biol* 1999, **9:**R554-557.
329. Kilzer JM, Stracker T, Beitzel B, Meek K, Weitzman M, Bushman FD: **Roles of host cell factors in circularization of retroviral dna.** *Virology* 2003, **314:**460-467.
330. Ott DE: **Purification of HIV-1 virions by subtilisin digestion or CD45 immunoaffinity depletion for biochemical studies.** *Methods Mol Biol* 2009, **485:**15-25.
331. Klinger PP, Schubert U: **The ubiquitin-proteasome system in HIV replication: potential targets for antiretroviral therapy.** *Expert Rev Anti Infect Ther* 2005, **3:**61-79.
332. Le Rouzic E, Morel M, Ayinde D, Belaidouni N, Letienne J, Transy C, Margottin-Goguet F: **Assembly with the Cul4A-DDB1DCAF1 ubiquitin ligase protects HIV-1 Vpr from proteasomal degradation.** *J Biol Chem* 2008, **283:**21686-21692.
333. Douglas JL, Viswanathan K, McCarroll MN, Gustin JK, Fruh K, Moses AV: **Vpu directs the degradation of the human immunodeficiency virus restriction factor BST-2/Tetherin via a {beta}TrCP-dependent mechanism.** *J Virol* 2009, **83:**7931-7947.
334. Conticello SG, Harris RS, Neuberger MS: **The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G.** *Curr Biol* 2003, **13:**2009-2013.
335. Sadowski M, Sarcevic B: **Mechanisms of mono- and poly-ubiquitination: Ubiquitination specificity depends on compatibility between the E2 catalytic core and amino acid residues proximal to the lysine.** *Cell Div* 2010, **5:**19.
336. Downs JA, Jackson SP: **Involvement of DNA end-binding protein Ku in Ty element retrotransposition.** *Mol Cell Biol* 1999, **19:**6260-6268.
337. Sawada M, Sun W, Hayes P, Leskov K, Boothman DA, Matsuyama S: **Ku70 suppresses the apoptotic translocation of Bax to mitochondria.** *Nat Cell Biol* 2003, **5:**320-329.
338. Gu Y, Jin S, Gao Y, Weaver DT, Alt FW: **Ku70-deficient embryonic stem cells have increased ionizing radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination.** *Proc Natl Acad Sci U S A* 1997, **94:**8076-8081.
339. Nussenzweig A, Chen C, da Costa Soares V, Sanchez M, Sokol K, Nussenzweig MC, Li GC: **Requirement for Ku80 in growth and immunoglobulin V(D)J recombination.** *Nature* 1996, **382:**551-555.
340. Daniel R, Litwin S, Katz RA, Skalka AM: **Computational analysis of retrovirus-induced scid cell death.** *J Virol* 2001, **75:**3121-3128.
341. Ott DE, Coren LV, Copeland TD, Kane BP, Johnson DG, Sowder RC, 2nd, Yoshinaka Y, Oroszlan S, Arthur LO, Henderson LE: **Ubiquitin is covalently attached to the p6Gag proteins of human immunodeficiency virus type 1 and simian immunodeficiency virus and to the p12Gag protein of Moloney murine leukemia virus.** *J Virol* 1998, **72:**2962-2968.
342. Pornillos O, Garrus JE, Sundquist WI: **Mechanisms of enveloped RNA virus budding.** *Trends Cell Biol* 2002, **12:**569-579.

343. Schubert U, Ott DE, Chertova EN, Welker R, Tessmer U, Princiotta MF, Bennink JR, Krausslich HG, Yewdell JW: **Proteasome inhibition interferes with gag polyprotein processing, release, and maturation of HIV-1 and HIV-2.** *Proc Natl Acad Sci U S A* 2000, **97:**13057-13062.
344. Sampson DA, Wang M, Matunis MJ: **The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification.** *J Biol Chem* 2001, **276:**21664-21669.
345. Bernier-Villamor V, Sampson DA, Matunis MJ, Lima CD: **Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1.** *Cell* 2002, **108:**345-356.
346. Bueno MT, Garcia-Rivera JA, Kugelman JR, Morales E, Rosas-Acosta G, Llano M: **SUMOylation of the lens epithelium-derived growth factor/p75 attenuates its transcriptional activity on the heat shock protein 27 promoter.** *J Mol Biol* 2010, **399:**221-239.
347. Lin DY, Huang YS, Jeng JC, Kuo HY, Chang CC, Chao TT, Ho CC, Chen YC, Lin TP, Fang HI, et al: **Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors.** *Mol Cell* 2006, **24:**341-354.
348. Cuchet-Lourenco D, Boutell C, Lukashchuk V, Grant K, Sykes A, Murray J, Orr A, Everett RD: **SUMO pathway dependent recruitment of cellular repressors to herpes simplex virus type 1 genomes.** *PLoS Pathog* 2011, **7:**e1002123.
349. Knipscheer P, Flotho A, Klug H, Olsen JV, van Dijk WJ, Fish A, Johnson ES, Mann M, Sixma TK, Pichler A: **Ubc9 sumoylation regulates SUMO target discrimination.** *Mol Cell* 2008, **31:**371-382.
350. Tatham MH, Kim S, Jaffray E, Song J, Chen Y, Hay RT: **Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralog selection.** *Nat Struct Mol Biol* 2005, **12:**67-74.
351. Zhu J, Zhu S, Guzzo CM, Ellis NA, Sung KS, Choi CY, Matunis MJ: **Small ubiquitin-related modifier (SUMO) binding determines substrate recognition and paralog-selective SUMO modification.** *J Biol Chem* 2008, **283:**29405-29415.
352. Kolesar P, Sarangi P, Altmannova V, Zhao X, Krejci L: **Dual roles of the SUMO-interacting motif in the regulation of Srs2 sumoylation.** *Nucleic Acids Res* 2012.
353. Lin D, Tatham MH, Yu B, Kim S, Hay RT, Chen Y: **Identification of a substrate recognition site on Ubc9.** *J Biol Chem* 2002, **277:**21740-21748.
354. Ao Z, Jayappa KD, Wang B, Zheng Y, Wang X, Peng J, Yao X: **Contribution of host nucleoporin 62 in HIV-1 integrase chromatin association and viral DNA integration.** *J Biol Chem* 2012, **287:**10544-10555.
355. Kalpana GV, Reicin A, Cheng GS, Sorin M, Paik S, Goff SP: **Isolation and characterization of an oligomerization-negative mutant of HIV-1 integrase.** *Virology* 1999, **259:**274-285.
356. Pichler A, Gast A, Seeler JS, Dejean A, Melchior F: **The nucleoporin RanBP2 has SUMO1 E3 ligase activity.** *Cell* 2002, **108:**109-120.

357. Saitoh H, Sparrow DB, Shiomi T, Pu RT, Nishimoto T, Mohun TJ, Dasso M: **Ubc9p and the conjugation of SUMO-1 to RanGAP1 and RanBP2.** *Curr Biol* 1998, **8:**121-124.
358. Dawlaty MM, Malureanu L, Jeganathan KB, Kao E, Sustmann C, Tahk S, Shuai K, Grosschedl R, van Deursen JM: **Resolution of sister centromeres requires RanBP2-mediated SUMOylation of topoisomerase IIalpha.** *Cell* 2008, **133:**103-115.
359. Kirsh O, Seeler JS, Pichler A, Gast A, Muller S, Miska E, Mathieu M, Harel-Bellan A, Kouzarides T, Melchior F, Dejean A: **The SUMO E3 ligase RanBP2 promotes modification of the HDAC4 deacetylase.** *Embo J* 2002, **21:**2682-2691.
360. Miyauchi Y, Yogosawa S, Honda R, Nishida T, Yasuda H: **Sumoylation of Mdm2 by protein inhibitor of activated STAT (PIAS) and RanBP2 enzymes.** *J Biol Chem* 2002, **277:**50131-50136.
361. Desfarges S, Salin B, Calmels C, Andreola ML, Parissi V, Fournier M: **HIV-1 integrase trafficking in S. cerevisiae: a useful model to dissect the microtubule network involvement of viral protein nuclear import.** *Yeast* 2009, **26:**39-54.
362. Rain JC, Cribier A, Gerard A, Emiliani S, Benarous R: **Yeast two-hybrid detection of integrase-host factor interactions.** *Methods* 2009, **47:**291-297.
363. Maertens G, Vercammen J, Debyser Z, Engelborghs Y: **Measuring protein-protein interactions inside living cells using single color fluorescence correlation spectroscopy. Application to human immunodeficiency virus type 1 integrase and LEDGF/p75.** *Faseb J* 2005, **19:**1039-1041.
364. Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, Christ F, Debyser Z: **The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding.** *J Biol Chem* 2005, **280:**17841-17847.
365. Astiazaran P, Bueno MT, Morales E, Kugelman JR, Garcia-Rivera JA, Llano M: **HIV-1 integrase modulates the interaction of the HIV-1 cellular cofactor LEDGF/p75 with chromatin.** *Retrovirology* 2011, **8:**27.
366. Hendrix J, Gijsbers R, De Rijck J, Voet A, Hotta J, McNeely M, Hofkens J, Debyser Z, Engelborghs Y: **The transcriptional co-activator LEDGF/p75 displays a dynamic scan-and-lock mechanism for chromatin tethering.** *Nucleic Acids Res* 2011, **39:**1310-1325.
367. Parvez MKU: **Thesis titled: Identification and characterization of new cellular interacting proteins of HIV-1 integrase.** 2011.
368. Fink LS, Lerner CA, Torres PF, Sell C: **Ku80 facilitates chromatin binding of the telomere binding protein, TRF2.** *Cell Cycle* 2010, **9:**3798-3806.
369. Fattah FJ, Lichter NF, Fattah KR, Oh S, Hendrickson EA: **Ku70, an essential gene, modulates the frequency of rAAV-mediated gene targeting in human somatic cells.** *Proc Natl Acad Sci U S A* 2008, **105:**8703-8708.
370. Daugaard M, Baude A, Fugger K, Povlsen LK, Beck H, Sorensen CS, Petersen NH, Sorensen PH, Lukas C, Bartek J, et al: **LEDGF (p75) promotes DNA-end resection and homologous recombination.** *Nat Struct Mol Biol* 2012, **19:**803-810.

371. Rothkamm K, Kruger I, Thompson LH, Lobrich M: **Pathways of DNA double-strand break repair during the mammalian cell cycle.** *Mol Cell Biol* 2003, **23:**5706-5715.
372. Boutell C, Cuchet-Lourenco D, Vanni E, Orr A, Glass M, McFarlane S, Everett RD: **A viral ubiquitin ligase has substrate preferential SUMO targeted ubiquitin ligase activity that counteracts intrinsic antiviral defence.** *PLoS Pathog 2011*, **7:**e1002245.
373. Hicks C, Gulick RM: **Raltegravir: the first HIV type 1 integrase inhibitor.** *Clin Infect Dis* 2009, **48:**931-939.