Virus-host interactions in early HIV-1 infection

Joanna Ceinwen Rowley

Thesis submitted for the degree of Doctorate of Philosophy

University College London

February 2015

I, Joanna Ceinwen Rowley, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

__Acknowledgements__

First and foremost I would like to thank Greg for his supervision, his patience, his encyclopedic knowledge, and his relentless enthusiasm for 'finding out stuff'.

I also sincerely thank Dave Selwood, Justin Warne, and their fellow chemists for their contribution to this project. I thank Dan Depledge for teaching me everything I know about SureSelect and Ale Riccombeni for handling all things bioinformatics.

I thank Adam and Tan for mentoring me when I first started in the lab and answering my endless questions. I thank Sarah, Petra, and Jane for celebrating and commiserating with me every step of the way. I thank Laura, Jane, and Rich for being an amazing support network – I couldn't have done it without you! I thank everyone else in the Towers lab and in wing 3.3 for being the most intelligent and entertaining group of people I have ever had the fortune of meeting.

I thank all my friends for putting up with me talking about viruses and for making sure I have good reasons to leave the lab! In particular I thank Tom, for keeping me sane, and Izzy, for bringing the fun.

Last, but most definitely not least, I wish to thank my parents for their unwavering love and support.

Abstract

After entry into a target cell, HIV-1 must traverse the cytoplasm and enter the nucleus in order to integrate into host chromatin. Until recently, the viral CA core was thought to play a passive role in infection, simply delivering its contents to the cytoplasm soon after entry. However, CA has recently been found to interact with and/or determine interaction with several nuclear entry cofactors. In this study, we investigate the consequences of CA interacting with the host mRNA-processing factor CPSF6. We provide evidence that the interaction between CA and cytoplasmic CPSF6 during early infection dictates the downstream nuclear entry pathway and integration site selection of HIV-1, despite not affecting viral titre in cell lines. We identify clinically relevant CTL escape mutations that lie within the CPSF6 binding pocket of CA and alter the relationship of HIV-1 with the CPSF6-dependent nuclear entry pathway. The phenotypes of these mutants suggest that conservation of this pathway is important *in vivo* but also that the pathway differs between cell types. In addition, we examine the antiviral mechanisms of C-terminal truncations of CPSF6 and the small-molecule CPSF6 peptidomimetics PF74 and BI-1. We design a series of compounds predicted to interact with the CPSF6-binding pocket of CA and screen them for antiviral activity as well as the ability to alter the nuclear entry pathway of HIV-1. Of particular interest are a compound that exhibits a very different inhibition profile from PF74/BI-1 and a compound that causes the virus to use a CPSF6-independent nuclear entry pathway without reducing viral titre. Together, the findings in this study confirm and characterise the role of CPSF6 as an HIV-1 cofactor and demonstrate that drugs targeting the CA-CPSF6 interaction could inhibit the virus through multiple different mechanisms.

Contents

List of figures

List of tables

List of abbreviations

6HB	6-helix bundle
Ab	antibody
AIDS	acquired immune deficiency syndrome
AP-1	activator protein 1
AP-2	activating protein 2
APC	antigen presenting cell
ApC	aphidicolin
APS	ammonium persulfate
ART	antiretroviral therpay
ARV	AIDS Related Virus
ARV	antiretroviral
AU	approximately unbiased
AZT	azidothymidine
BIV	Bovine Immunodeficiency Virus
BLAST	Basic Local Alignment Search Tool
BP	bootstrap probability
BSA	bovine serum albumin
BST-2	bone marrow stromal cell antigen 2
CA	capsid
cART	combined ART
CC_{50}	cytotoxic concentration 50%
CCL	chemokine ligand
CCR5	C-C chemokine receptor 5
CD4	cluster of differentiation 4
CD4bs	CD4 binding site
CD8	cluster of differentiation 8
CDK	cyclin-dependent kinase
CFIm	cleavage factor Im
cGAMP	cyclic GMP-AMP or cyclic [G(2'-5')pA(3'-5')p]
cGAS	cyclic GMP-AMP synthase
ChIP	chromatin IP

CHR	C-terminal helical region
CLR	C-type lectin receptors
CMV	Cytomegalovirus
cPPT	central PPT
CPSF	cleavage and polyadenylation specificity factor
CPSF6	cleavage and polyadenylation specificity factor 6
CsA	cyclosporine A
Ct	cycle threshold
CTD	C-terminal domain
CTL	cytotoxic T lymphocyte
CXCR4	C-X-C chemokine receptor 4
CycT1	cyclin T1
Cyp	cyclophilin
CypA	cyclophilin A
DBP	DNA binding protein
DC	dendritic cell
DC-SIGN	dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DHS	DNase I hypersensitivity site
DIG	digoxigenin
DIS	dimer initiation signal
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulphoxide
DNA-PK	DNA-dependent protein kinase
DRB	5,6-Dichloro-1-β-$_D$-ribofuranosylbenzimidazole
DSIF	DRB sensitivity-inducing factor
EC_{50}	effective concentration 50%
EDTA	ethylenediaminetetraacetic acid
eGFP	eGFP
EIAV	Equine Infectious Anaemia Virus
ELISA	enzyme-linked immunosorbent assay
EM	electron microscopy
Env	evelope

ER	endoplasmic reticulum
ERAD	ER-associated degradation
ESCRT	endosomal sorting complexes required for transport
EU	5-ethynyl uridine
FAM	6-carboxyfluorescein
FBS	foetal bovine serum
FISH	fluorescent *in situ* hybridisation
FIV	Feline Immunodeficiency Virus
FlAsH	fluorescin arsenical helix binder
FoC	fate of capsid
Fos	FBJ murine osteosarcoma viral oncogene homolog
FRAP	fluorescence recovery after photobleaching
FSC-H	forward scatter height
G418	Geneticin G-418 sulphate
Gag	group-specific antigen
GALT	gut associated lymphoid tissue
GFP	green fluorescent protein
gp120	glycoprotein 120kDa
gp160	glycoprotein 160kDa
gp41	glycoprotein 41kDa
GPI	glycophosphatidylinositol
GWAS	genome-wide association study
HA	hemagglutinin
HAART	highly active ART
Hck	haemopoietic cell kinase
HIV	Human Immunodeficiency Virus
HLA	human leukocyte antigen
HP1	heterochromatin protein 1
HRP	horse radish peroxidase
HSPG	heparan sulphate proteoglycans
HTLV	Human T-Lymphotropic Virus
IBD	IN binding domain
ICAM-1	intracellular adhesion molecule 1

IDAV	Immune Deficiency Associated Virus
IDU	injecting drug user
IF	immunofluorescence
IFN	interferon
IFNAR2	IFN-α/β receptor 2
IL-2	interleukin 2
IN	integrase
INI	IN inhibitor
InSTI	integration strand transfer inhibitor
ISG	IFN-stimulated gene
IU	infectious unit
Jun	Jun proto-oncogene
LAV	Lymphadenopathy Associated Virus
LB	lysogeny broth
LEDGF/p75	lens epithelium-derived growth factor
LEDGIN	LEDGF inhibitor
LINE	long interspersed element
LTNP	long-term non-progressor
LTR	long terminal repeat
MA	matrix
MDM	monocyte-derived macrophage
MFI	mean fluorescence intensity
MHC	major histocompatibility complex
MLV	Murine Leukaemia Virus
MOI	multiplicity of infection
MoMLV	Moloney MLV
MVB	multivesicular body
MVV	Maedi-Visna Virus
MxB	Myxovirus resistance protein B
NC	nucleocapsid
Nef	negative regulatory factor
NELF-E	negative elongation factor E
NES	nuclear export signal

NFAT	nuclear factor of activated T-cells
NFκB	nuclear factor κB
NHEJ	non-homologous end joining
NHR	N-terminal helical region
NLS	nuclear localisation signal
NMR	nuclear magnetic resonance
NPC	nuclear pore complex
NTD	N-terminal domain
Nudix	nucleoside diphosphate linked to x
Nup	nucleoporin
Nup153	nucleoporin 153kDa
Nup155	nucleoporin 155kDa
Nup214	nucleoporin 214kDa
Nup358	nucleoporin 358kDa, a.k.a. RANBP
Nup85	nucleoporin 85kDa
Nup98	nucleoporin 98kDa
NVP	nevirapine
NXF1	nuclear RNA export factor 1
OWM	Old World monkey
p-TEFb	positive transcription elongation factor b
PAGE	polyacrylamide gel electrophoresis
PAK2	p21-activated kinase 2
PAMP	pathogen-associated molecular pattern
PAP	poly(A) polymerase
PAS	poly(A) signal
PB	polybrene
PBL	peripheral blood lymphocyte
PBMC	peripherial blood mononucleocytes
PBS	phosphate buffered saline
PBS	primer binding site
PCP	*Pneumocystis carinii* pneumonia
PEG	polyethylene glycol
PFA	paraformaldehyde

PFV	Prototype Foamy Virus
PHA	phytohemagglutinin
PI	PR inhibitor
PIC	pre-integration complex
POD	peroxidase
PPIase	peptidyl-prolyl isomerase
PPT	polypurine tract
PR	protease
PRR	pattern recognition receptor
PVDF	polyvinylidene difluoride
qPCR	quantitative/real time PCR
RANBP2	Ran binding protein 2, a.k.a. Nup358
Rev	regulator of expression of virion proteins
RLU	relative light units
RNAPII	RNA polymerase II
RPMI	Roswell Park Memorial Institute Medium
RRE	Rev responsive element
RRM	RNA recognition motif
RS	arginine/serine-rich
RT	reverse transcriptase
RT-qPCR	reverse transcription qPCR
RTC	reverse transcription complex
SAMHD1	sterile α motif and HD domain-containing protein 1
SC35	splicing component 35kDa
SDM	site directed mutagenesis
SDS	sodium dodecyl sulphate
SGA	single genome analysis
SIN	self-inactivating
SINE	small interspersed element
SIV	Simian Immunodeficiency Virus
SIVcpz*Ptt*	SIV chimpanzee *Pan troglodytes troglodytes*
SIVgor	SIV gorilla
SIVmac	SIV rhesus macaque

SIVmne	SIV Macaca nemestrina
SIVsm	SIV sooty mangebey
SP	slow progressor
SP	spacer peptide
SR	serine/arginine-rich
SSC-H	side scatter height
STAT3	signal transducer and activator of transcription 3
STING	stimulator of IFN genes
SV40	Simian Virus 40
TAE	Tris acetate EDTA
TAK-1	tumour growth factor-β activated kinase 1
TAMRA	tetramethylrhodamine
TAP	transporter associated with antigen presentation
TAR	transactivation responsive region
Tat	trans-activator of transcription
TBS	Tris buffered saline
TCR	T cell receptor
TE	Tris EDTA
TEM	transmission EM
TEMED	tetramethylethylenediamine
TFP	transframe protein
TI	therapeutic index
TI	transcriptional interference
T_m	melting temperature
TNFα	tumour necrosis factor α
TNPO3	transportin 3, a.k.a. TRN-SR2
TRIM	tripartite motif
TRN-SR2	transportin of serine/arginine proteins 2, a.k.a. TNPO3
UTR	untranslated region
V3	variable loop 3
Vif	viral infectivity factor
VLP	virus-like particle
Vpr	viral protein r

Vpu	viral protein u
Vpx	viral protein x
VS	virological synapse
VSV	Vesicular Stomatitis Virus
VSV-G	VSV G envelope protein
WHO	World Health Organisation
WT	wild-type
YFP	yellow fluorescent protein
Ψ	HIV-1 packaging signal

1 Chapter 1. Introduction

1.1 HIV-1 - the causative agent of AIDS

1.1.1 HIV-1 – a primate Lentivirus

The Baltimore classification system places the family *Retroviridae* in Class VI, a group defined by a positive single-stranded RNA (+ssRNA) genome that replicates through a double-stranded DNA (dsDNA) intermediate by virtue of a virally encoded RNA-dependent DNA polymerase, reverse transcriptase (1-3). The *Retroviridae* are further characterised by a dimeric genome of 7-10Kbp, virions 80-100nm in diameter, a lipid envelope modified with *Env*-encoded viral glycoproteins, and an integrase enzyme that catalyses permanent integration of the dsDNA reverse transcripts into a host genome forming a provirus (4). Retroviruses all encode two polypeptides – Gag and Pol – which encode for the viral structural and enzymatic proteins including *Pro*, which encodes the viral protease that cleaves these polypeptides (5, 6).

The family *Retroviridae* comprises two subfamilies, *Spumavirinae* and *Orthoretrovirinae*, the latter of which includes the genus *Lentivirus* (4). Retroviruses can also be categorised into simple and complex viruses, with Lentiviruses being considered complex (7). Lentiviruses were originally characterised by their slow ("lente" in Latin) progress from infection to disease. Five serogroups of *Lentivirus* have now been identified; equine, bovine, feline, ovine, and primate, the latter comprising the species Simian Immunodeficiency Virus (SIV) and Human Immunodeficiency Virus 1 and 2 (HIV-1 and 2). In addition to *Gag, Pol,* and *Env*, Lentiviruses are characterised by the presence of regulatory genes *Tat* and *Rev*, species-specific accessory genes (8), and the unique ability among retroviruses to replicate efficiently in non-dividing cells (9).

1.1.2 The discovery of HIV-1 as the causative agents of AIDS

Patients with symptoms of an acquired immune deficiency syndrome (AIDS) were first officially documented in 1981 when a cohort of homosexual men in

the USA presented with cancers such as Kaposi's sarcoma and opportunistic infections such as *Pneumocystis carinii* pneumonia (PCP) (10-12), which are characteristic of immune deficiency, before eventually succumbing to the disease (13).

In 1983, the laboratory of Luc Montagnier isolated a novel retrovirus from a lymph node of a homosexual male with multiple lympadenopathies, which they believed to be a causative agent of AIDS, and subsequently named the virus Lymphadenopathy Associated Virus (LAV). (14). They identified it as a member of the *Retrovirirdae* by its density and reverse transcriptase activity. Its similarity to the *Deltaretrovirus* Human T-Lymphotropic Virus I (HTLV-I) with respect to T-cell tropism was noted, but this new virus differed from HTLV-I in antigenicity and the two viruses shared only 10% genetic homology (14). In 1984, the laboratory of Robert Gallo reported the isolation of another virus thought to be a causative agent of AIDS, which they named HTLV-III (15). Unlike LAV, the HTLV-III isolate replicated efficiently in T-cell lines and thus testing of this virus in a laboratory became possible. It was later confirmed that Gallo's isolate originated in the laboratory of Montagnier (16) and that this virus was the same species as other reported AIDS-associated isolates such as AIDS Related Virus (ARV) and Immune Deficiency Associated Virus (IDAV), and thus the name HIV was coined to encompass all isolates (17). HIV was identified as a *Lentivirus* by genetic similarity to Maedi-Visna Virus (MVV), an ovine *Lentivirus* (18, 19). In 1986 a virus was isolated that shared many core epitopes with HIV but differed significantly in its envelope glycoproteins (20). This virus was subsequently named HIV-2 and HIV renamed HIV-1.

1.1.3 The HIV/AIDS pandemic

Within a year of discovery, AIDS was no longer confined to males, homosexuals, or the USA. Injecting drug users (IDUs) (21) and haemophiliacs (22) had been identified as high-risk groups, and it was understood that heterosexual and mother-to-child transmission of the pathogen was possible (23). Today, sexual transmission is by far the most common route of infection, but mother-to-child

transmission (*in utero*, during delivery, and through breast-feeding) is also a common route in less developed countries. Despite the first cases of AIDS being diagnosed in the USA, the origins of HIV-1 have been traced to sub-Saharan Africa. From here the virus has become pandemic, but 70% of HIV-1 infected people are still found in this region. The HIV prevalence rate in sub-Saharan Africa is estimated at 5%, with no other region of the world reaching above 1% (24). As of December 2013, UNAIDS reported that 35 million people are infected with HIV-1 globally. The number of new HIV-1 infections per year peaked in 1997 and has fallen by 21% since, but there are still an estimated 2.1 million new infections every year (25).

1.1.4   The clinical course of HIV-1 infection and AIDS

The stage of HIV-1 infection is generally assessed by measuring the level of viraemia in a patients plasma and their CD4+ T cell count. The typical clinical course of infection for someone not receiving any antiretroviral treatment in depicted in figure 1 (26). For 1-2 weeks after initial infection with HIV-1 there is an asymptomatic 'eclipse' phase during which the virus is replicating and disseminating from the site of infection but cannot be detected (26). HIV-1 RNA becomes detectable by day 11 (on average), antigen by day 16, and antibodies by day 22, at which point the patient is said to have seroconverted (25, 27). The virus now replicates rapidly in both the gut associated lymphoid tissue (GALT) and in peripheral lymphoid tissues and the viral load increases exponentially (28). During this 'acute' phase, many patients experience a febrile illness including lymphadenopathy. Coinciding with peak viraemia, an adaptive immune response against the virus appears, which is followed by a rapid drop in viremia (26) until a consistent viral load (known as a 'set-point') is established. The subsequently 'chronic' stage of HIV-1 infection lasts for 8-10 years (in the absence of treatment) and is often referred to as clinical latency due to the lack of symptoms. However, the virus itself is not latent but is replicating rapidly. As a result, CD4$^+$ T cells, which are the primary target cells of HIV-1, are being extremely rapidly turned over. Over this period, viral load gradually increases and CD4$^+$ T cell count gradually decreases. Once the CD4$^+$ T
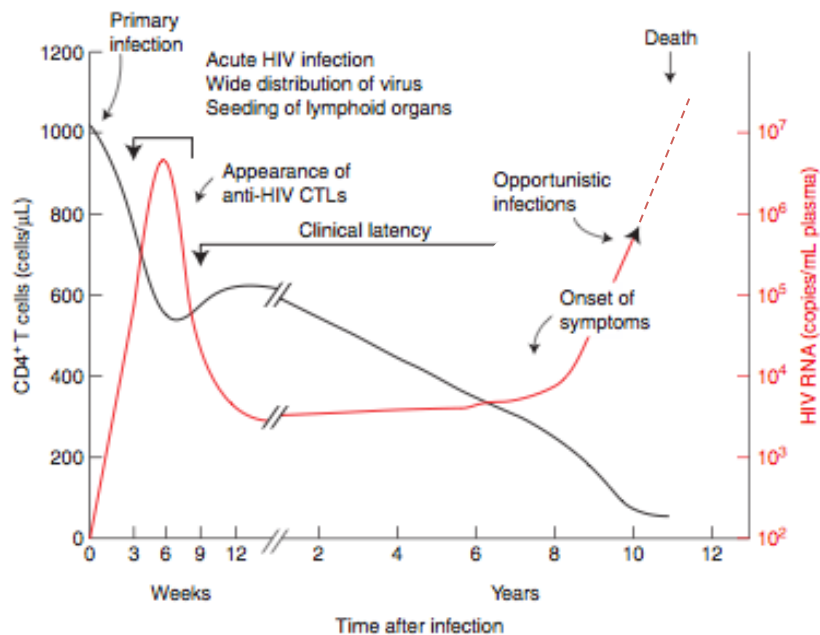
Figure 1. The clinical course of HIV-1 infection
The typical course of an untreated HIV-1 infection over time (weeks or years post-infection, where a double forward slash indicates a change in units) assessed by CD4+ T cell count (cells/µl blood) in black or HIV RNA levels (copies/ml plasma) in red. Arrows highlight significant clinical events/phases. Adapted from Coffin and Swanstrom 2013

cell count becomes too low, the immune system is unable to control opportunistic infections such as PCP, extrapulmonary TB, and Cryptococcosis. AIDS is defined by the World Health Organisation (WHO) as an HIV-1 infection accompanied by at least one of the following; a CD4+ T cell count below 200 cells/μl, a CD4$^+$ T cell percentage of total lymphocytes of less than 15%, or any one of a list of AIDS-associated conditions, which includes the opportunistic infections described above as well as HIV-associated encephalopathy and tumours (26).

1.1.5 Treatment of HIV-1

The number of people dying from AIDS per year peaked in 2005 at 2.3million and by 2013 this number had dropped by 35%, primarily due to access to antiretroviral therapy (ART), which are now received by 37% of those in need. The WHO recommends treatment be initiated once a patient's CD4+ T cell count falls below 500 cells/μL. Five different classes of antiretroviral (ARV) that target four different stages of replication are currently licensed. HIV-1 evolves extremely rapidly and so it readily develops resistant to single drug. Therefore, combined/highly-active antiretroviral therapies (cART/HAART) involving three different drugs from at least two different classes are used. Although these treatments can effectively inhibit viral replication and spread and the immune system can eliminate already productively infected cells, neither these drugs nor the immune system can eliminate the reservoir of latent proviruses. Therefore, the virus rebounds rapidly after therapy is stopped (29). The need for life-long therapy, the side effects of therapy, and the emergence of multi-resistant variants has meant that a great deal of research is being carried out into new classes of ARVs, cure strategies, and into vaccines and other prophylactics.

1.1.6 The origins of HIV-1

Since the discovery of HIV-1, extensive phylogenetic studies have sought to identify the origin/s of the virus. HIV-1 comprises four lineages, namely group M (major group), O (outlier group), N (non-O/M), and P. Group M is the only

pandemic strain, accounting for the vast majority of HIV infections worldwide, and can be further subcategorised into several subtypes. Group O represents less than 1% of HIV infections and is found primarily in Cameroon and Gabon whilst only 13 cases of group N and two cases of group P have been identified thus far (30, 31). HIV-2 is quite distantly related to HIV-1 and is largely restricted to western Africa. Patients infected with HIV-2 tend to have significantly lower viral loads than those infected with HIV-1, do not transmit the virus as efficiently, and rarely progress to AIDS (32). HIV-2 comprises 8 lineages, A-H, although only groups A, B, and F have proven human-human transmission (30). Phylogenetic studies have revealed that each HIV group arose from a separate zoonotic transmission event; as shown in phylogenetic trees in figure 2, HIV-1 group M and N are most closely related to the chimpanzee SIVs (SIVcpz*Ptt*), HIV-1 group P is most closely related to gorilla SIVs (SIVgor), and all HIV-2 groups to sooty mangebey SIVs (SIVsm) (30). Genetic molecular clocks date the origin of HIV-1 group M to the very beginning of the twentieth century (33) and HIV-2 groups A and B to 1940-50 (34). Although the route of transmission between non-human primate (ape or monkey) and human for all of the aforementioned viruses remains debated, hunting/eating bushmeat is considered a likely culprit.
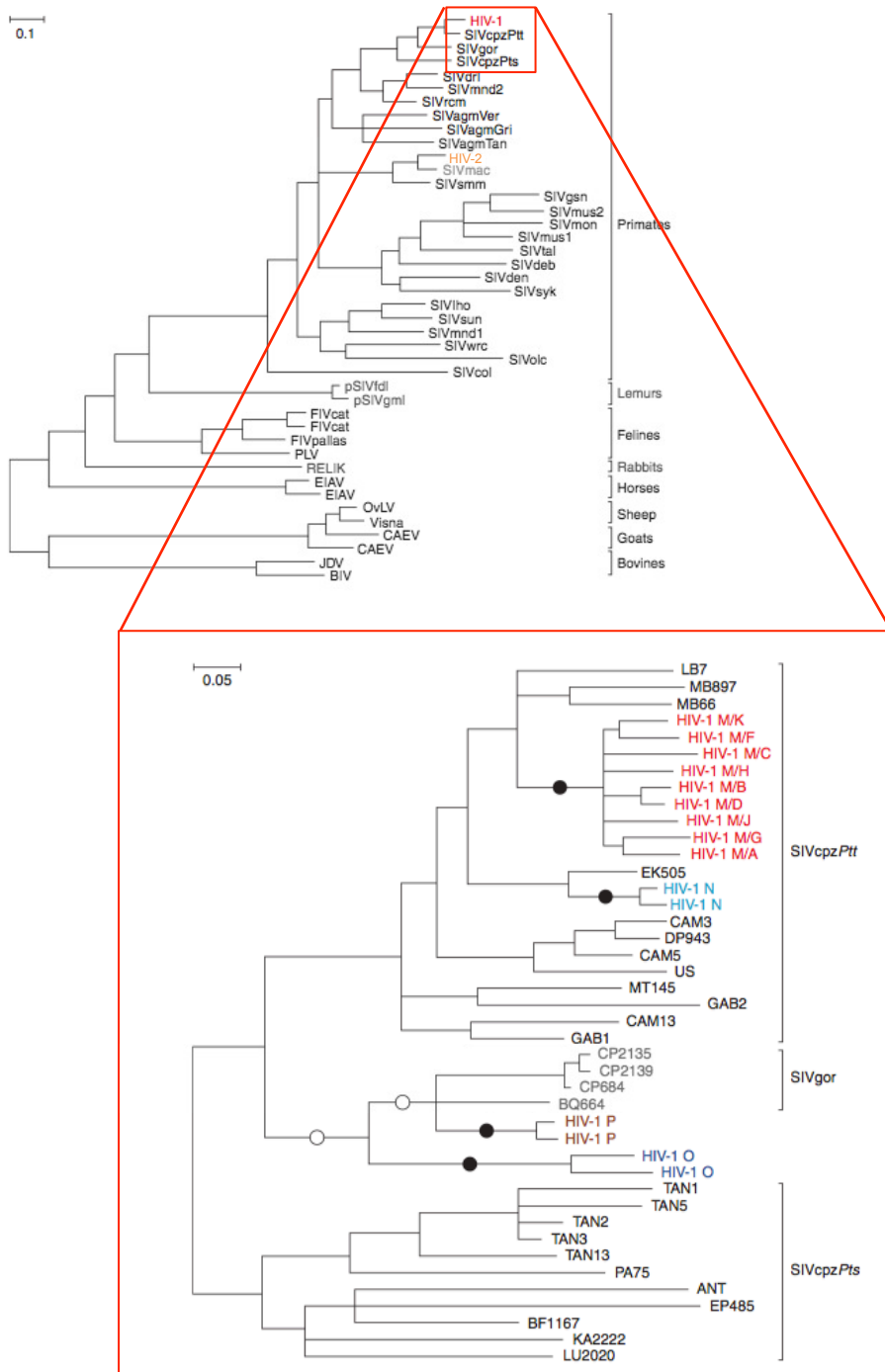
Figure 2. The phylogenetic origins of HIV
Phylogenetic trees showing the relationship between HIV-1 and other Lentiviruses (top). Red boxes/lines indicate a magnified region, which shows the relationship of HIV-1 groups, SIVcpz, and SIVgor (bottom). HIV-2 is shown in orange, HIV-1 M group in red, N group in cyan, P group in brown, and O group in blue. Filled black circles indicate a zoonotic transmission event. Open circles indicate two alternative possible zoonotic transmissions events. The scale bar shows the branch length that represents 0.1 nucleotide substitutions per site (10% divergence). Adapted from Sharp et al. 2011.

1.2 The lifecycle of HIV-1

1.2.1 The genomic structure of HIV-1

Figure 3 shows the structure of the 9.7Kbp HIV-1 genome (35). The HIV-1 *gag* gene encodes the group-specific antigen (Gag, p55) polypeptide that is cleaved into the structural proteins matrix (MA, p17), capsid (CA, p24), and nucleocapsid (NC, p7), as well as p6, and the spacer peptides 1 and 2 (SP1/2). Approximately 1 in 20 times that *gag* is translated, a ribosomal frameshift occurs meaning that *pol* is also translated, thus producing a Gag-Pol fusion protein (36). *Pol* encodes the enzymes protease (PR, p10), reverse transcriptase (RT, p51/66), and integrase (IN, p31). The *env* gene encodes the envelope glycoprotein of 160kDa (gp160), which is subsequently cleaved by the host protease furin into two glycoproteins of 120 and 41kDa (gp120 and gp41), which form a stable heterodimer, Env. The HIV-1 genome also encodes the regulatory proteins, trans-activator of transcription (Tat) and regulator of expression of virion proteins (Rev), and accessory proteins, viral infectivity factor (Vif), viral protein r (Vpr), viral protein u (Vpu), and negative regulatory factor (Nef). The coding sequence of HIV-1 is flanked by two non-coding long terminal repeats (LTRs), which comprise many transcription factor binding sites and are also essential for integration of viral DNA into the host genome.

1.2.2 The structure of the mature HIV-1 virion

Figure 4 shows the most complete model of the mature HIV-1 virion thus far (37). It shows that the virion's envelope is studded with gp120/gp41 trimers, as well as host proteins usually found on the surface of cells such as intracellular adhesion molecule 1 (ICAM-1) (38). Associated with and directly beneath this lipid bilayer is a shell of MA (39). Beneath MA, Vpr and p6 can be found, and beneath that the viral core. The core is conical in shape and is composed of hexamers and pentamers of CA (detailed in section 1.2.7). Inside this core, lies the dimeric +ssRNA genome associated with the viral NC and Vif as well as enzymes required for the early stages of HIV-1 infection; IN and RT (40).

F R A M E

1

2

3

0   1000   2000   3000   4000   5000   6000   7000   8000   9000

5' LTR

p17   p24   p2 p1   p7 p6
gag

prot   p51 RT   p15 RNase   p31 int
pol

vif

vpr

vpu

gp120   gp41
env

tat

rev

nef
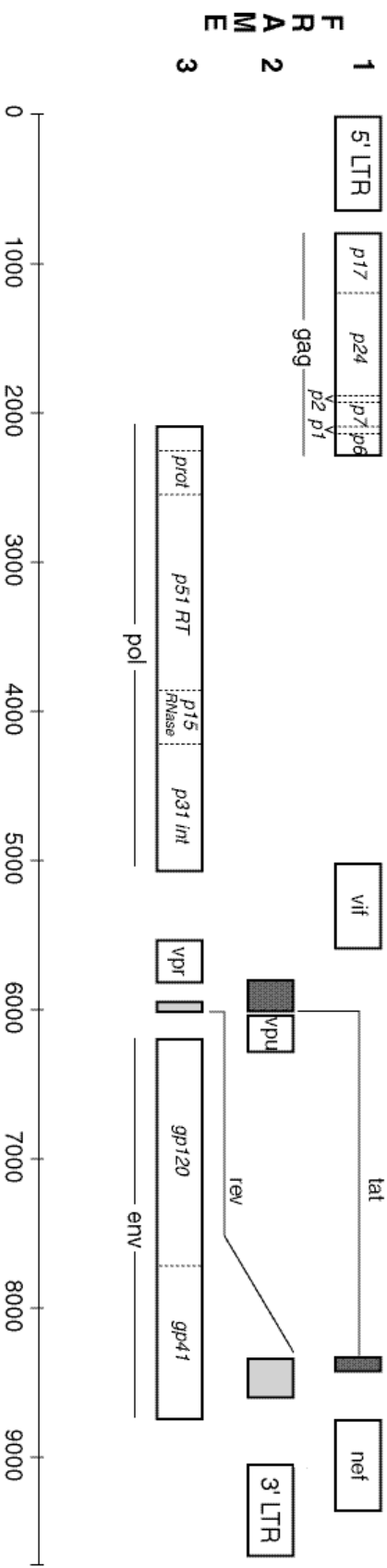
3' LTR

Figure 3. The HIV-1 genome
The layout of the HIV-1 genome based upon HXB2, a laboratory-adapted strain often used as a reference strain. Axis shows nucleotide number. Boxes indicate ORFs (except for the 5' and 3' LTRs, which are not protein coding, but comprise many important regulatory elements). Vertical positioning of ORFs shows which reading frame (as labeled on the left) each gene is read in. *gag*, *pol*, and *env* encode polyproteins that are cleaved into multiple discrete proteins during maturation, as indicated by the dashed lines within the boxes. The *tat* and *rev* spliced exons are shaded and linked by lines. Image modified from Los Alamos National Laboratory HIV-1 Gene Map 2014.

33
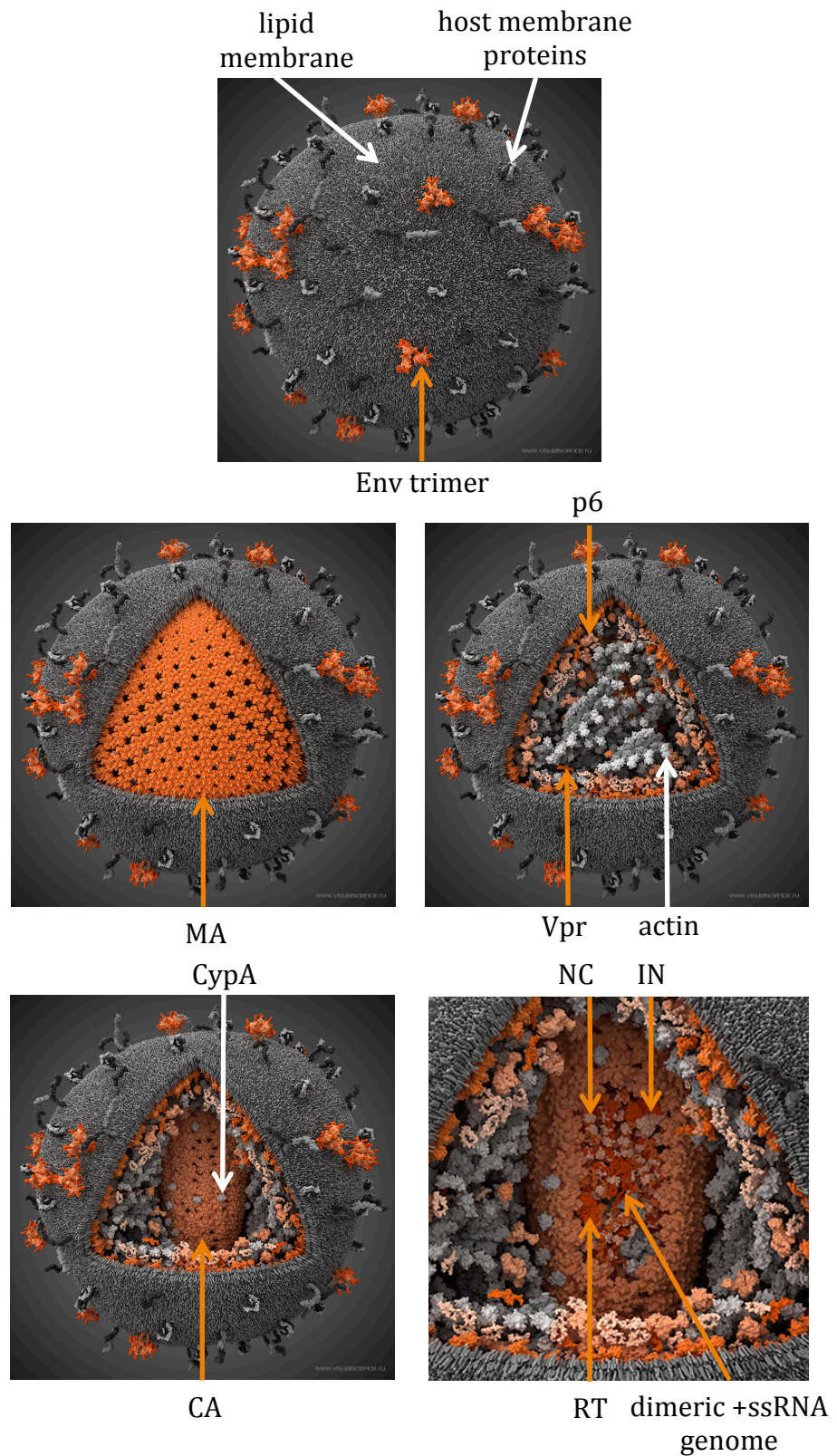
Figure 4. The structure of the mature HIV-1 virion

An artists representation of mature HIV-1 based upon data taken from 33 scientific publications. Host-derived membranes/proteins in greys, viral proteins in orange/ red. Host membrane proteins include ICAM-1, HLA-II, and CD55. Adapted from images by Konstantinov. I for Visual Science 2011.

1.2.3   An overview of the HIV-1 lifecycle

Free HIV-1 virions attach to target cells when gp120/gp41 trimers embedded in their membrane engage with their receptor, cluster of differentiation 4 (CD4), and a coreceptor, either C-C chemokine receptor 5 (CCR5) or C-X-C chemokine receptor 4 (CXCR4). This leads to fusion of the viral and cellular membranes, such that the viral core is delivered to the cytoplasm (41). Prior to nuclear entry the viral genome undergoes reverse transcription, which converts the +ssRNA genome into an integration competent dsDNA, and the virus sheds its CA core in a process called uncoating (40). The resulting complex then enters the nucleus via the nuclear pore. In the nucleus, IN orchestrates integration of the viral DNA into the host genome, forming a provirus (42). The provirus is then transcribed by the cellular RNA polymerase II. mRNA encoding the early proteins Tat, Rev, and Nef is exported and translated by normal cellular processes. Tat and Rev then regulate transcription, export, and translation of all other viral proteins (36). Viral proteins and newly synthesised genomic +ssRNA assemble at the cell membrane and the virus buds from the cell surface. The viral protease, which is packaged within this immature virion, then cleaves the Gag and Gag-Pol polypeptides resulting in maturation of the particle (43). This infectious particle is then able to infect another target cell.

1.2.4   Entry of HIV-1 into target cells

Attachment and entry of HIV-1 into target cells is mediated by the viral protein Env. The *Env* gene encodes gp160, which forms key disulphide bonds and is heavily glycosylated in the Golgi apparatus (41). It is subsequently cleaved by the host protease furin, forming heterodimers of gp120/gp41 often simply referred to as Env, and is secreted to the cell surface where it forms trimers. Gp41 comprises a cytoplasmic C-terminal tail, a transmembrane domain, and an ectodomain consisting of two large helical regions (the C- and N-terminal helical regions, CHR and NHR) separated by a loop region and followed by a short fusion peptide domain (38, 44). In contrast, gp120 is a globular protein that lies on the cell surface and consists of five conserved loops (C1-5) and five highly variable loops (V1-5) (41). Env trimers are incorporated onto the surface

of new HIV-1 virions during budding (detailed in section 1.2.11.1) such that each virion has an estimated 10-15 Env 'spikes' on their surface (38). Env is the only viral protein exposed on the surface of virions and is thus a major target for antibodies. Antibody recognition of Env is reduced by its glycosylation and the variability of its most exposed regions (38, 41). It is not known how many spikes are required for HIV-1 entry but some reports' estimates are as low as one or two (38). Maintaining a low number of viral proteins on the virion surface is thought to further contribute to antibody evasion.

Virions can reversibly attach to the surface of cells non-specifically, through interaction of Env with heparan sulphate proteoglycans (HSPGs) or of two non-viral proteins such a ICAM-1 and LFA, or specifically, for example interaction of Env with C-type lectin receptors (CLRs) such as dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN) (figure 5A) (45). Although this attachment step is not essential and the virus can proceed directly to receptor engagement, attachment is thought to help bring the virus in closer proximity to the cell and thus enhance the probability of Env interacting with receptor (41).

The receptor for HIV-1 Env is CD4, a member of the immunoglobulin superfamily that enhances T cell receptor (TCR) signaling (46, 47). Direct interaction between the highly conserved CD4 binding site (CD4bs) of gp120 and CD4 is the first irreversible step in HIV-1 entry (figure 5B) (48). This interaction causes dramatic conformational changes in gp120 leading to exposure of variable loop 3 (V3) and formation of a 4-stranded β-sheet called the bridging sheet, both of which are coreceptor binding sites (figure 5C) (38). HIV-1 can use two different coreceptors, CXCR4 (49) and CCR5 (50-54), both of which are seven-transmembrane domain G-coupled chemokine receptors. Different HIV-1 strains utilise different coreceptors; strains using CCR5 are named R5 viruses, those using CXCR4 are named X4 viruses, and those able to use both are name R5X4 viruses. Coreceptor binding induces further conformational changes in gp120/gp41 such that the hydrophobic fusion peptide of gp41 is inserted into the host cell membrane and the fusion peptide
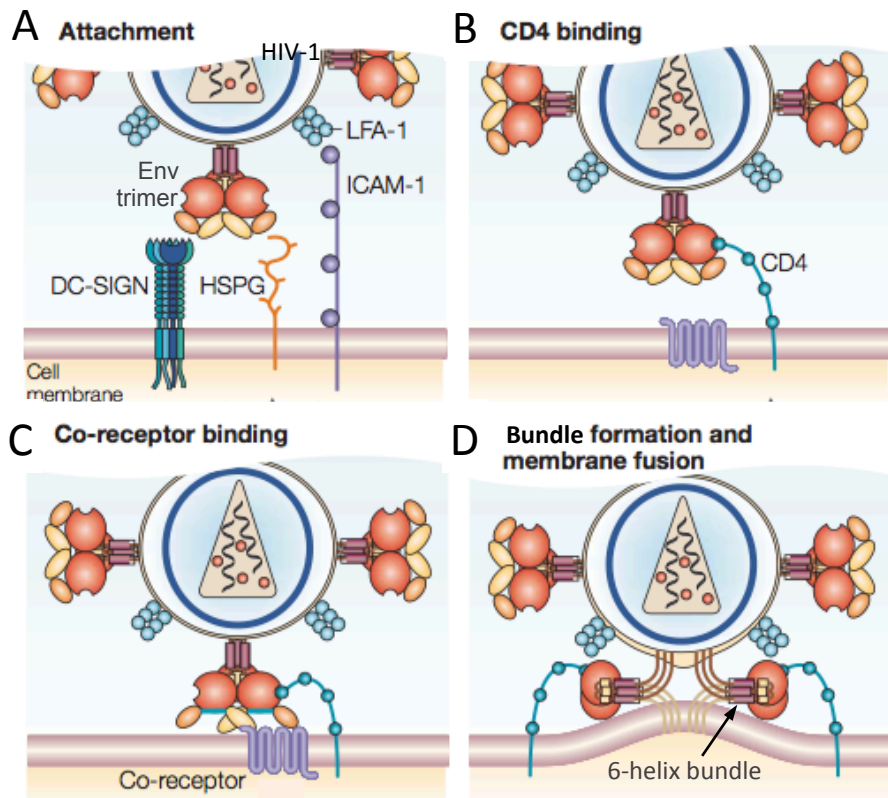
Figure 5. HIV-1 attachment and entry
(A) Attachment is mediated by weak, non-essential, reversible interactions between proteins within the viral membrane and those within the target cell membrane. For example, Env and DC-SIGN, Env and HSPGs, or LFA-1 and ICAM-1. **(B)** Interaction between Env and CD4 is irreversible. **(C)** Env interaction with CD4 causes conformational changes that reveal regions of Env that interact with one of the HIV-1 co-receptors, CXCR4 or CCR5. **(D)** Interaction between Env and one of the co-receptors results in further conformational changes and the formation of a 6-helix bundle, which brings the viral membrane into such close contact with the target cell plasma membrane that they are able to fuse. Adapted from Shattock and Moore 2003.

folds at a hinge region, bringing its CHR and NHR together. In the context of an Env trimer this forms a six-helix bundle (6HB) (55) which brings the viral and cellular membranes into very close contact such that they fuse (figure 5D). Fusion of these membranes creates a fusion pore, through which the contents of the virion can enter the cytoplasm (56).

The vast majority of transmitted viruses are found to be R5 viruses (57). The importance of CCR5 as a coreceptor is demonstrated by the fact that homozygosity for *ccr5Δ32* – a 32bp deletion in the *ccr5* gene which causes retention of CCR5 in the ER – confers a high level of resistance to HIV-1 (41). In around half of patients the virus population undergoes a 'coreceptor switch' and becomes predominantly X4 virus. This usually occurs late in infection and is associated with progression to AIDS (58). The reasons why HIV-1 preferentially uses one coreceptor over another at different stages of infection remain poorly understood. It has been proposed that X4 viruses are more easily recognised by the immune system and so do not arise until significant immune dysregulation has occurred, late in HIV-1 infection. It has also been hypothesised that the virus is forced to switch to use CXCR4 because the body becomes depleted of CCR5-expressing cells and the virus needs to expand/alter its tropism (58).

Another unresolved debate about HIV-1 entry is whether it occurs at the cell surface, in endosomes, or whether it can occur via both routes. Until recently, the observations that HIV-1 entry does not require a low pH nor endocytosis of CD4 and that expression of Env is sufficient to cause cell fusion was thought to support the hypothesis that entry occurred at the cell surface (38). However, a recent study by Miyauchi, et al. (59) used single-virion fluorescence imaging to separate the stages of lipid-mixing (fusion of lipid membranes) and contents-mixing (entry of NC fused to green-fluorescent protein (GFP) into the cytoplasm) and found that although lipid-mixing could occur at the cell surface, contents-mixing was delayed and occurred from inside endosomes in a dynamin-dependent manner. Intriguingly, contents-mixing could occur very close to the nucleus, suggesting that the viral core would only have to traverse a

short distance of cytoplasm. Importantly, this study was carried out exclusively in HeLa TZM-Bl cells, and the route of entry may well be cell-type dependent.

As well as spread via cell-free virus HIV-1 can spread directly cell-to-cell. This occurs when an infected cell comes into contact with an uninfected cell and forms a virological synapse (VS). The CD4-gp120 interaction is essential for the formation of a VS, but it is thought to be stabilised by integrin-ICAM-1 interactions. There is evidence that the infected cell polarises such that its secretory machinery is oriented such that virions assemble at the VS (60). Cell-to-cell spread is estimated to be between 10 and 100 times more efficient than cell-free spread *in vitro*. This is likely due several factors including a higher multiplicity of infection. Cell-to-cell spread is also thought to be an important method of evading antibody recognition *in vivo* (61).

Two HIV-1 entry inhibitors have been licensed thus far. Enfuvirtide is a 36-mer peptide that mimics part of the gp41 CHR and so binds the NHR, preventing formation of the 6HB (62). Unfortunately, this drug has low oral bioavailability and so must be injected (41). Maraviroc inhibits the use of CCR5 by binding to this host protein and allosterically altering its gp120 binding site (63). It was predicted that resistance to CCR5 inhibitors would arise quickly because it would simply select for outgrowth of X4 variants. However, this only occurs in a minority of patients, which argues for the notion that X4 viruses have some replicative disadvantage *in vivo* (64).

1.2.5 Trafficking of HIV-1 in early infection

Once the viral core enters the host cell cytoplasm it must traffic toward the nucleus and it is therefore unsurprising that a large number of cytoskeletal components have been implicated in HIV-1 infection. McDonald et al. (65) used live cell imaging to show that cytoplasmic virions labeled with GFP-Vpr travel along microtubules and that inhibitors of the retrograde microtubule transport motor dynein causes accumulation of HIV-1 virions at the cell periphery. In addition, Arhel, et al. (66) used live imaging to show that IN-labeled virions

move toward the nucleus in a manner that is kinetically characteristic of both microtubules and actin but that their movement slows down as they near the NPC, indicative of movement via perinuclear actin filaments called microfilaments. There has been difficulty in examining many purported relationships between HIV-1 and the cytoskeleton due to the toxicity of inhibiting/depleting cytoskeletal/motor proteins, the apparent redundancy in the system, and the fact that the composition of the virus whilst trafficking is debatable and may change over time (particularly with respect to the state of the CA core as detailed in section 1.2.7) (67).

1.2.6   Reverse transcription

Reverse transcription is the process by which DNA is generated from an RNA template. The RT enzyme encoded by HIV-1 not only uses the +ssRNA viral genome as a template to produce a complementary -ssDNA, but also degrades the +ssRNA template and subsequently synthesises the +ssDNA creating a dsDNA product. This requires RT to have RNA and DNA-dependent DNA polymerase activity as well as RNase H activity.

Pol encodes a polypeptide that is cleaved from its precursor into the functional peptides p9, p66, and p31 during virion maturation (detailed in section 1.2.11.3). p66 comprises both an RNA/DNA-dependent DNA polymerase domain and an RNase H domain. p66 initially forms homodimers but one monomer is subsequently cleaved into p51, thus losing its RNase H domain. This leads to the formation of functional p66/p51 heterodimers, often referred to simply as RT (68). As shown in figure 6, p66 forms the active sites of both the polymerase and the RNase, whilst p51 performs a structural role. The structure of the polymerase domain of p66 is often described as a right hand, with the 'thumb', 'fingers', and 'palm' designated as labeled in figure 6. Unliganded, the thumb nearly meets the fingers, but the thumb moves back to reveal the nucleic acid binding cleft and the fingers move forward to form a dNTP binding pocket (40).
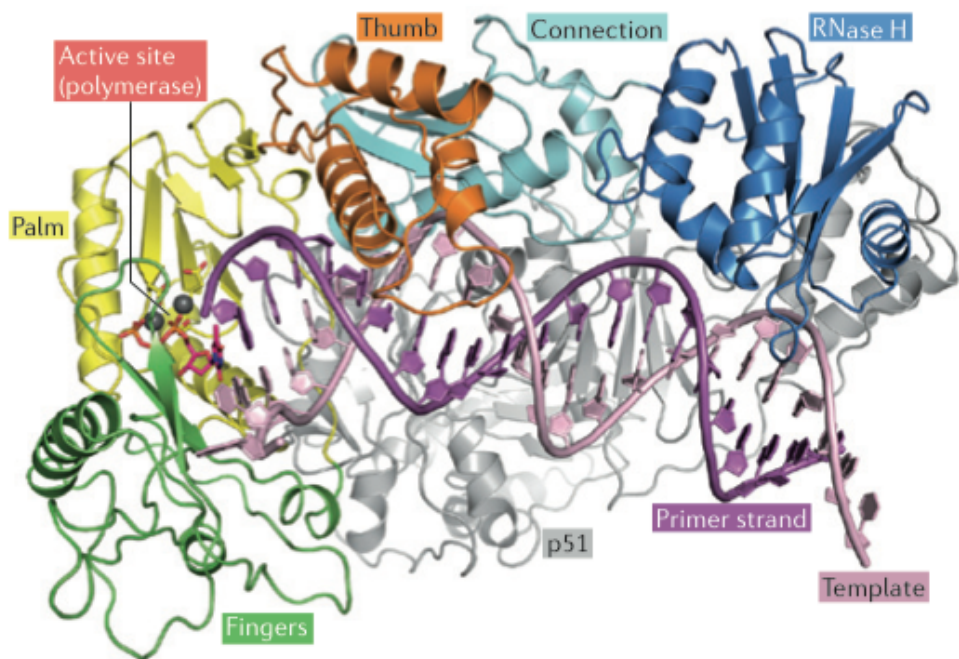
Figure 6. The structure of HIV-1 reverse transcriptase
Crystal structure of HIV-1 reverse transcriptase (p55/p66) in complex with DNA (PDB 1RTD from Huang et al. 1998). The inactive scaffolding p51 subunit is in grey, whilst the enzymatically p66 subunit is coloured according to the labels. The DNA template strand and primer strand are highlighted in purples. A dTTP molecules in the polymerase active site is in red. The two Mg2+ ions in the polymerase active site are shown as grey spheres. Three domains of p66 are labeled by analogy to the human hand (fingers, palm, thumb). Image taken from Engelman and Cherepanov 2012.

The trigger of initiation of reverse transcription is unknown, but the viral core lattice has 8-10nm gaps between hexamers that are large enough to allow entry of dNTPs (and the RT inhibitors described below) (69, 70). It is therefore possible that exposure to dNTPs inside the target cell is sufficient to trigger reverse transcription (71). All retroviruses use tRNAs as primers for their RT enzymes. HIV-1 uses tRNA^{Lys3}, which is packaged into HIV-1 virions via interaction with Gag and Pol, respectively (72). The 3' sequence of tRNA^{Lys3} is complementary to an 18nt long sequence within the HIV-1 genome called the PBS (primer binding site), which lies immediately 3' to the 5' LTR (40) (figure 7, steps 1 and 2). Priming from this tRNA^{Lys3}, RT synthesises minus strand DNA from the plus strand RNA template (figure 7, step 3). The distance between the polymerase and the RNase H domain of RT can accommodate 17 to 18 nucleotides and thus RNase H mediated degradation begins 17-18nts behind DNA synthesis (73). R is a direct repeat that is present at both the 5' and 3' ends of the genome. Having reverse transcribed and then degraded the template's 5' R, the minus strand DNA is free to transfer (or 'jump') and to anneal to the template's 3' R (figure 7, step 4). RT can now continue to synthesise the minus strand DNA along the length of the genome and to degrade the RNA template as it goes. HIV-1 has two poly purine tracts (PPTs), which are resistant to RNase H mediated degradation (figure 7, step 5). The PPT lies at the 3' end of the genome whilst the cPPT (central PPT) lies within *pol*. These PPTs can now acts as the primer for synthesis of the plus strand DNA (figure 7, step 6). The cPPT is not essential for HIV-1 replication, but is thought to enhance the efficiency of plus strand DNA synthesis. As the tRNA^{Lys3} is still attached to the 5' end of the minus strand DNA RT also reverse transcribes the first 18 nucleotides of tRNA into plus strand DNA until it reaches a modified nucleotide that it cannot read and RNase H activity subsequently removes the tRNA (figure 7, step 7). These overhanging 18 nucleotides of tRNA are complementary to the PBS, and so the plus strand DNA transfers/jumps to anneal to the PBS of the minus strand DNA (figure 7, step 8). DNA synthesis now proceeds in both directions to complete the plus and minus strands and RNase H removes the PPT RNA primers, creating dsDNA (40) (figure 7, step 9). It is important to note that this dsDNA product is longer than the +ssRNA template; both the 5' and 3' end now have

1. RNA template

2. Minus strand DNA priming

3. Minus strand DNA synthesis 1

4. Minus strand transfer

5. Minus strand DNA synthesis 2

6. Plus strand DNA priming

7. Plus strand DNA synthesis 1

8. Plus strand transfer

9. Plus strand DNA synthesis 2

R U5 PBS *gag* *pol* cPPT *env* PPT U3 R

tRNA^{Lys3}

PBS *gag* *pol* cPPT *env* PPT U3 R U5
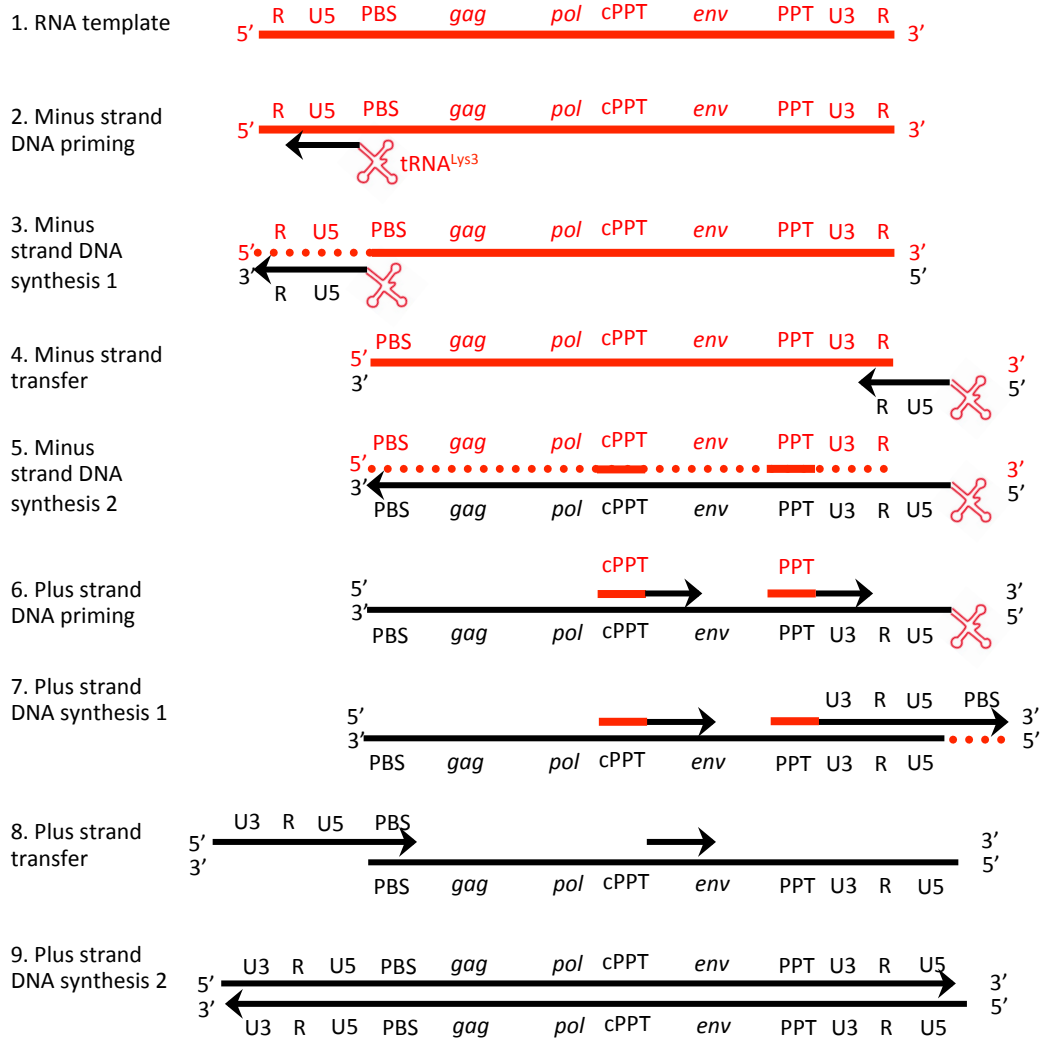
Figure 7. HIV-1 reverse transcription
The process of HIV-1 reverse transcription in 9 stages. RNA is shown in red, DNA in black. U3 is the 3' UTR, U5 is the 5' UTR, R is the repeat region – all are regions of the LTR. PBS is the primer binding site. PPT is the polypurine tract. cPPT is central PPT. Arrows show direction of DNA polymerisation. Dotted line signifies degradation by RNase H, which follows ~18nt behind polymerisation.

the same LTR consisting of U3, R, and U5. This is required for integration of this DNA into the host genome (detailed in section 1.2.9.1). HIV-1 evolves rapidly and, until recently, this was largely attributed to the fact that RT has no proofreading function and thus has very low template fidelity. However, the human RNA polymerase II also has no proofreading ability, meaning that mutations can be introduced during both reverse transcription of incoming genomes and during synthesis of new genomes. Although it is impossible to measure the error rate of these two enzymes individually in the context of an HIV-1 infection, the overall mutation rate per HIV-1 replication cycle is 2×10^{-5}/nucleotide and thus the error rate of RT can be no higher than this. Reverse transcription further contributes to viral diversity by allowing recombination between the two viral genomes to occur. Each HIV-1 virion contains two +ssRNA genomes and RT readily moves between them, particularly during long pauses, the minus and plus strand transfers, or when it encounters nicks in the genomic RNA (40).

When reverse transcribing, the viral complex is often referred to as the reverse transcription complex (RTC) but there is much debate about the composition of this complex and the roles of its constituent proteins. NC is a nucleic acid chaperone that coats HIV-1 genomic RNA. It has been implicated in genome dimerisation and packaging, $tRNA^{Lys3}$ unfolding, annealing of $tRNA^{Lys3}$ to the genome, facilitating strand transfers, and destabilising secondary structures in RNA such that RT can progress (74). IN is also present in the RTC and, curiously, many IN mutations can inhibit reverse transcription. However, the enzymatic activity of IN is not required for reverse transcription to proceed and thus it is thought that IN plays a structural role in the RTC (75). There is particular disagreement as to whether CA is present in the RTC. Some studies have suggested that the majority of CA is lost very early after infection, others that it is lost sequentially during reverse transcription, and others that it remains an intact core until immediately prior to nuclear entry (details of uncoating are discussed in section 1.2.7). Other proteins such as Vpr, Vif, and Tat are purportedly part of the RTC but their roles have not yet been elucidated.

Two classes of ARVs inhibit reverse transcription; nucleoside and non-nucleoside reverse transcriptase inhibitors (NRTIs and NNRTIs). NRTIs mimic dNTPs and are incorporated into viral DNA by RT, but they lack the 3' OH group required for incorporation of the next dNTP and are thus chain terminators (76, 77). In contrast, NNRTIs are allosteric inhibitors that cause significant conformational changes to both the polymerase active site and the primer binding site of RT (76). The NRTI azidothymidine (AZT) was the first licensed ARV and RT inhibitors remain the most widely used ARVs, with 11 of the 29 licensed ARVs being NRTIs and 5 being NNRTIS.

1.2.7 The HIV-1 core and uncoating

As discussed in section 1.2.4, binding and fusion is followed by delivery of the viral core to the cytoplasm. Pornillos, et al. (78) recently published the most accurate atomic level model of the HIV-1 core thus far. Their model, shown in figure 8A, is a fullerene core composed of 1056 CA subunits, which comprise an NTD and CTD that fold independently and are connected by a flexible linker. These CA monomers form a lattice of 166 hexamers and 12 pentamers (depicted in figure 8B). These pentamers induce curvature to the hexameric lattice and thus allow it to close. The conical shape of the core is due to the asymmetrical distribution of pentamers, five at one end and seven at the other. Whilst interactions between N-terminal domains (NTDs) form the hexamers/pentamers, interactions between C-terminal domains (CTDs) connect these hexamers/pentamers into a cone.

After delivery into the target cell cytoplasm, the virus must reverse transcribe and then enter the nucleus to integrate. Many studies categorise incoming cytoplasmic viral complexes as either reverse transcription complexes (RTCs) or pre-integration complexes (PICs). Attempts to determine and quantify the components of these complexes have produced inconsistent results and many now consider it futile to attempt to define a complex that is likely constantly changing with regards to location, composition (with regards to both viral and host proteins), and structure. We therefore use the terms RTC and PIC simply to
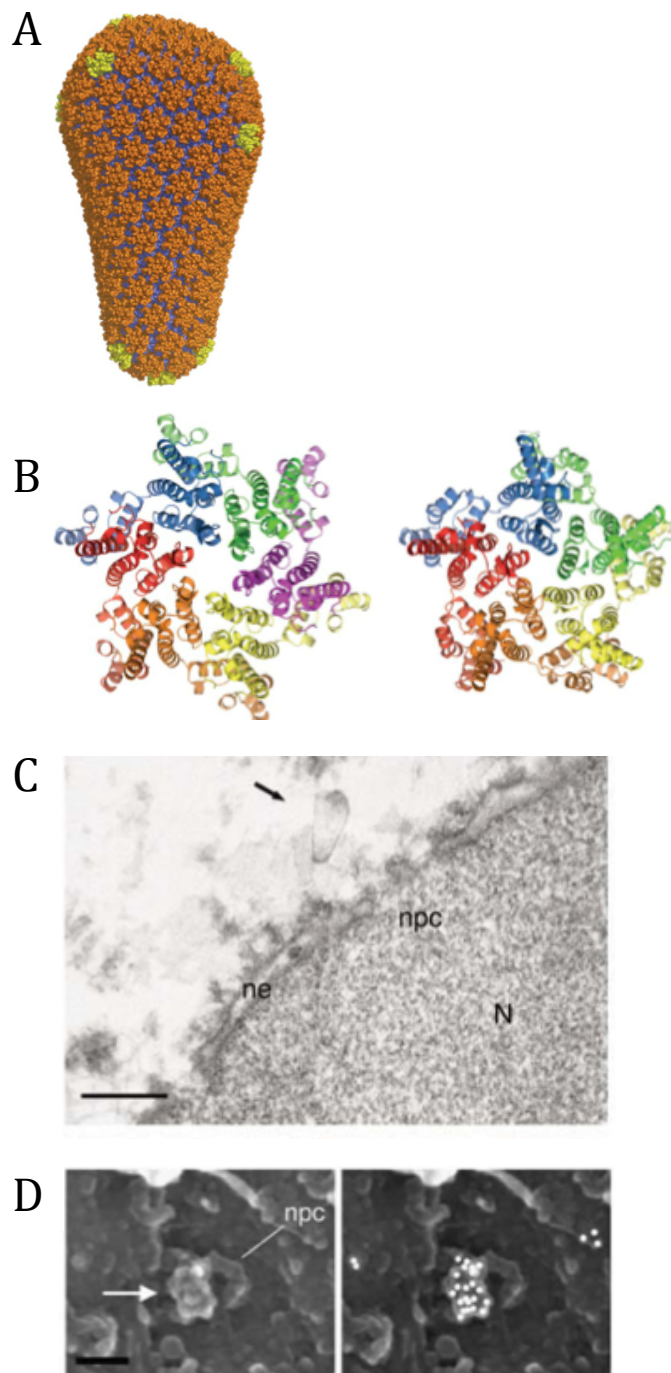
Figure 8. The structure of the mature HIV-1 core
(A) Model of the mature conical HIV-1 core. CTDs in purple, hexamer NTDs in orange, pentamer NTDs in yellow. **(B)** Crystal structures of CA hexamer and pentamer (PDB 3H47 and 3P05, respectively, both from Pornillos et al. 2011). Images in (A) and (B) all taken from Pornillos et al. 2011 **(C)** Cryo-scanning EM image of MT4 cells infected with HIV-1. Arrow highlights a conical core docking proximal to an NPC. N = nucleus. ne = nuclear envelope. npc = nuclear pore complex **(D)** EM images of P4 cells infected with HIV-1. Left panel shows EM image alone, whilst right panel has immunogold p24 labeling superimposed, revealing an HIV-1 core docking at an NPC. (C-D) Scale bars 100nm. Images taken from Arhel et al. 2007.

refer to complexes that are competent for reverse transcription and integration, respectively. HIV-1 can efficiently infect non-dividing cells and is as dependent upon nuclear pore proteins in dividing cells as in non-dividing cells, suggesting the virus enters the nucleus via the NPC regardless of cell cycle stage (79). It is agreed that the 50-60nm by 100-120nm core is too large to enter the nucleus via the 30-40nm wide nuclear pore complex (NPC) and so cores must disassemble in the cytoplasm prior to nuclear entry via a processing termed 'uncoating' (71). However, the location, kinetics, mechanism, and consequences of uncoating are highly controversial.

1.2.7.1 *The mechanism, timing, and location of uncoating*

Until recently, it was hypothesised that the viral core spontaneously uncoats soon after entry into the cytoplasm. Early work used transmission electron microscopy (TEM) to visualise virions during the early stages of infection and showed that the dense conical core was undetectable soon after entry (80). However, this was likely because cores are difficult to differentiate from other cellular structures unless they are cut along the correct axis to view their conical shape (71). Fractionation of infected cell lysates by sucrose gradient centrifugation was also used to suggest that CA dissociates from the viral genome within 1hr of infection (81, 82). However, this method has been heavily criticised for using strong detergents that disrupt important interactions.

The hypothesis that cores completely uncoat soon after entry has recently been challenged by reports that CA directly interacts with several nucleoporins and that these interactions are required for HIV-1 nuclear entry (discussed in section 1.5.1) (79, 83, 84). These observations revealed that at least some CA must remain associated with the virus until it reaches the NPC. One alternative hypothesis is that uncoating occurs gradually. This hypothesis is primarily supported by *in vitro* uncoating assays that suggest uncoating is biphasic. Virions are stripped of their membrane and MA layer with strong detergents. The resulting naked cores are then ultracentrifuged through sucrose such that the dense intact cores are pelleted but the less dense monomeric CA is not (85).

The proportion of intact CA is then assessed by western blot. This assay has been used to show that when cores are incubated at 37°C, they spontaneously uncoat in a biphasic manner; disassembly is rapid for 30 minutes before slowing down significantly, and is 90% complete after 3hrs (86). However, this technique yields highly variable results, uses harsh detergents that are known to induce uncoating, and does not allow uncoating to occur in a cellular environment. This latter argument is becoming increasingly important as more CA-binding host proteins are implicated in core uncoating (87-90). In order to overcome this problem, recent studies have used immunofluorescence (IF) and electron microscopy (EM) to visualise cytoplasmic viral complexes. Xu *et al.* showed that the RNA within RTCs was accessible to 5-ethynyl uridine (EU), a small molecule RNA dye, but not to RNase A (91). Whilst the authors interpreted this to mean that cores must partially open or uncoat to allow EU to reach the RNA, recent atomic-level structures of the HIV-1 core have shown that it is porous enough to allow entry of very small molecules such as EU and dNTPs without uncoating (78).

It has recently been discovered that cofactors interacting with CA are required to prevent HIV-1 DNA from being detected by cytoplasmic innate immune sensors (92, 93) (discussed further in section 3.3.7). The simplest mechanism as to how CA could be involved in hiding viral DNA from is by remaining intact throughout the cytoplasm and uncoating at the NPC. This is a particularly compelling hypothesis because several other viruses, including Herpes Simplex Virus I (94-96), some Adenoviruses (97, 98), and Granuloviruses (99), are known to dock at the NPC before delivering their genome into the nucleus. This hypothesis is supported by the visualisation of intact cores in the nuclear periphery. For example, McDonald *et al.* quantified the CA signal associated with RTCs (detected by incorporation of fluorescent dNTPS) and found it to be strong as that of extracellular virions (65), suggesting that cores are intact. Furthermore, these intact cores could be detected in perinuclear regions as early as 30mins post-infection (65). In agreement with these observations, Arhel, et al. (100) used cryo-scanning EM, immunogold p24 labeling, and DNA fluorescent *in situ* hybridisation (FISH) to show that intact cores containing

DNA can 'dock' at NPCs (figures 8C and D). They also showed that the disappearance of perinuclear p24 correlates with the appearance of nuclear DNA (100), which implies that these cores disassemble at the same time as their contents enters the nucleus. The development of extremely small tetracysteine tags (6 amino acids, 0.7kDa), which fluoresce when in complex with fluorescin arsenical helix binder (FlAsH), and super-resolution EM has allowed comparison of the shapes of IN-containing complexes both in extracellular virions and inside cells. As in extracellular virions, many cytoplasmic IN-containing complexes appeared as 100nm long conical structures and, importantly, were often seen in perinculear regions (101). Taken together, these findings have shown that intact conical HIV-1 cores containing DNA reverse transcripts can be found docked at NPCs.

Importantly, not all cytoplasmic viral complexes are intact and localised to the nuclear periphery. In fact, visualisation of individual cytoplasmic viral complexes has revealed that they differ hugely in both composition and location. For example, McDonald *et al.* show that only ~67% of RTCs contain detectable amounts of CA (65). Importantly, it has been calculated that only 1 in 8 virions that reverse transcribes will successfully integrate (102) and, as yet, there is no assay that can differentiate between viral complexes that would go on to be infectious and those that would not. It is therefore possible that the only infectious virions are those that remain intact until they reach the NPC. Although we cannot be sure as to which RTC composition represents infectious virions, the fact that intact cores have been visualised at the NPC is an excellent proof of principle.

1.2.7.2 The relationship between reverse transcription and uncoating
A complex relationship exists between reverse transcription and uncoating. There is a wealth of evidence suggesting that these two processes are interdependent and the consensus appears to be that the viral core must be in tact for reverse transcription to begin but that the growing reverse transcripts then drive uncoating. *In vitro* reverse transcription reactions have revealed that

RT often pauses for long periods of time (particularly during minus strand DNA synthesis) and has low processivity, frequently falling off its template (103). It is therefore hypothesised that a locally high concentration of RT around the RNA template must be maintained during reverse transcription to ensure completion. The most simple mechanism by which this could occur is that the majority of reverse transcription occurs within an intact core (71).

Evidence that reverse transcription drives uncoating is provided by the finding that inhibition of reverse transcription delays uncoating. Several studies have shown that inhibition of reverse transcription using nevirapine (NVP, an NNRTI) stabilises cores and slows/delays uncoating. This effect of NVP upon uncoating has been measured using three different assays. Firstly, the fate of capsid (FoC) assay is a biochemical assay that involves ultracentrifugation of virus-infected cell extracts through a sucrose gradient such that dense intact cores are pelleted but monomeric CA is not (104). This assay has been used to show that infection in the presence of NVP increases the amount of intact CA (105). Secondly, the effect of NVP upon uncoating has also been measured indirectly, as follows. TRIMCyp is an HIV-1 restriction factor (detailed in section 1.3.1) that only inhibits intact viral cores. It interacts with CA via its cyclophilin (Cyp) domain and so cyclosporine A (CsA), which also interacts with Cyp domains, can be used to prevent TRIMCyp from interacting with and restricting HIV-1. In order to measure the timing of uncoating, TRIMCyp-expressing cells are infected with HIV-1 in the presence of CsA (so restriction is prevented) and infection is allowed to proceed. CsA is then removed/washed out at various time points, allowing TRIMCyp to restrict the virus. When CsA is removed and TRIMCyp is still unable to restrict the virus, this indicates that the core has uncoated (106). This assay has been used to show that NVP slows uncoating (107). Thirdly, immunofluorescence has been used to show that NVP increases the proportion of cytoplasmic virions associated with CA (RTCs/PICs, labeled with GFP-Vpr) (100, 107). Together, these studies suggest that inhibition of reverse transcription prevents uncoating and therefore support the hypothesis that reverse transcription drives uncoating.

In light of these observations, it has been hypothesised that unstable cores may not be able to initiate reverse transcription and that hyperstable cores may be unable to complete reverse transcription. Forshey *et al.* used the *in vitro* uncoating assay described in section 1.2.7.1 to show that E45A(CA) uncoats more slowly than wild-type (WT) HIV-1 whilst P38A uncoats more quickly than WT HIV-1 (86). They interpreted this to mean that E45A cores are intrinsically hyperstable and that P38A cores are intrinsically unstable. They showed that both of these mutations are defective for reverse transcription and proposed that this was due to their stability defects. However, Yang *et al.* recently identified compensatory mutations for both E45A and P38A that rescued infectivity without altering stability (108). Furthermore, disruption of microtubules with nocodazole or with siRNA targeting the motors dynein or kinesin has been shown to delay uncoating without affecting reverse transcription (109). These data show that changes to core stability and/or the rate of uncoating do not necessarily inhibit reverse transcription.

Despite this, we cannot rule out the possibility that changes in stability may influence reverse transcription in some circumstances. Host proteins such as cyclophilin A (CypA) (89, 110) and small molecules such as PF74 (111) have been shown to affect reverse transcription by interacting directly with CA (detailed in section 1.5.2.2 and 5.1, respectively) and biochemical assays have been used to suggest that do so by altering core stability. However, these studies have produced inconsistent results (87, 90, 112, 113). An alternative hypothesis is that these molecules affect reverse transcription by altering the porosity of the HIV-1 core and the accessibility of dNTPs to the viral genome.

1.2.7.3 Host proteins influence uncoating

As well as CypA, other host proteins have been implicated in core stability and uncoating. Factors proposed to destabilise cores include transportin 3 (TNPO3) (87) and tripartite motif protein 5 (TRIM5) (114, 115), whilst factors proposed to stabilise cores include nucleoporin 153 (Nup153) (116) and Myxovirus resistance protein B (MxB) (117). The evidence supporting a role for these

proteins in uncoating is discussed in the relevant sections. The influence of host proteins upon uncoating may explain why the rate of uncoating is cell-type dependent (118).

1.2.8 HIV-1 nuclear entry

1.2.8.1 The nuclear pore complex

As discussed previously, Lentiviruses are characterised by their ability to efficiently infect non-dividing cells. In order to do this it must traverse the NPC, a 112MDa complex of at least 500 molecules of 30 different proteins that forms a 39nm diameter channel through which all nucleocytoplasmic trafficking occurs (119). The proteins comprising the NPC are named nucleoporins or Nups followed by a number denoting their molecular mass. Nucleoporins are classified into six categories according to their position and role within the NPC, as denoted in figure 9A. Molecules of up to 40kDa can diffuse through channels in the NPC, whereas larger molecules require active translocation (119). Many nucleoporins, particularly those lining the channel, contain extensive phenylalanine/glycine (FG) repeats, as shown in figure 9B (120). These flexible and dynamic repeat regions are thought to form a dense matrix or gel that acts as the pore's permeability barrier (121). All transportins/karyopherins identified thus far are at least mildly hydrophobic, which is thought to be essential for navigating through this selective channel (122). Karyopherins interact with their cargo via nuclear localisation/export signals (NLS/NESs), and their directionality is controlled by a gradient of RanGTP across the nuclear envelope. Interestingly, the NPC has been implicated in a range of activities aside from nucleocytoplasmic transport such as cell division, chromatin organisation, DNA repair, and gene expression (119).

1.2.8.2 Viral and host determinants of nuclear entry

As discussed earlier, HIV-1 must uncoat prior to nuclear entry due to the size constraints of the NPC. It is assumed that at least one component of the uncoated HIV-1 PIC must contain an NLS and many viral proteins have been

A

B
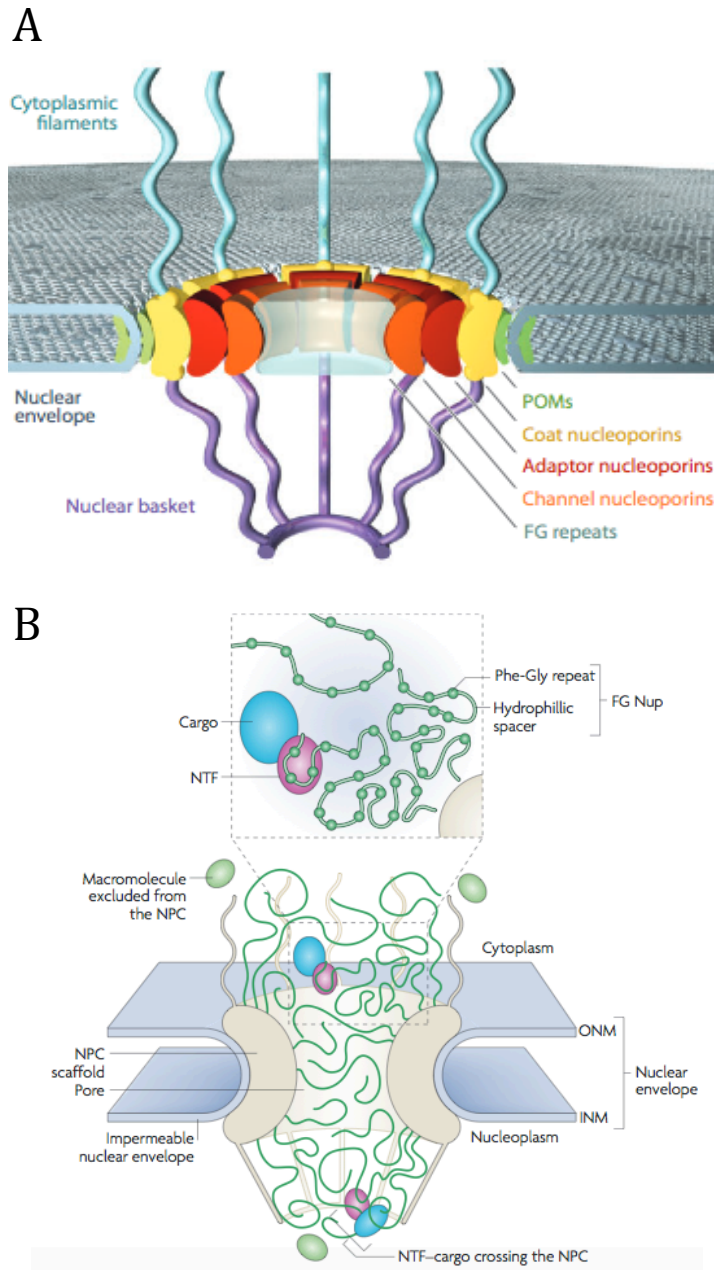
Figure 9. The structure of the nuclear pore complex
(A) A model of the nuclear pore complex. The four coloured concentric rings comprise four categories of nucleoporin: POMs (pore membrane proteins), and coat, adaptor, or channel nucleoporins. The cytoplasmic filaments and nuclear basket are two other categories of nucleoporin. Many of these nucleoporins contribute FG repeat regions to the NPC channel/pore. Image taken from Hoelz et al. 2011. **(B)** A model of the FG repeat organisation at the NPC. Green lines represent FG repeat regions of nucleoporins that line the NPC channel/pore, forming a matrix or filter through which cargo must travel in order to enter/exit the nucleus. NTF = nuclear transport factor. O/INM = outer/inner nuclear membrane. Image taken from Strambio-de-Castilla et al. 2010.

reported to contain them. Classical NLSs that can mediate nuclear entry when transferred to other proteins have been reported in MA, IN, and Vpr, and also in the cPPT. Whilst all of these factors have implicated in HIV-1 nuclear entry, subsequent studies have largely refuted these observations by identifying mutations in these regions that do not affect HIV-1 nuclear entry (9, 122). Yamashita and Emerman (123) showed that redundancy cannot explain these observations by making a chimeric HIV-1 containing MLV IN and mutating the proposed nuclear localisation determinants in MA, Vpr, and the cPPT, and showing that it is still able to infect non-dividing cells efficiently. However, this does not rule out the possibility of more redundant NLSs that have not yet been discovered.

More recently, CA has been identified as a major genetic determinant of nuclear entry. Although CA does not enter the nucleus itself, there is increasing evidence that it orchestrates many events prior to nuclear entry. It was found that chimeric HIV-1 with MLV CA is unable to infect non-dividing cells (124) and, similarly, that mutations in CA (such as A92E and G94D, which are discussed further in section 1.5.2.3) can cause HIV-1 to become cell-cycle dependent (125). As discussed in section 1.2.7, the virus must shed its CA core in order to enter the nucleus and aberrant uncoating can inhibit nuclear entry. CA has been found to interact with, and/or be the genetic determinant of, several nuclear entry cofactors. In doing so, CA may determine the route taken by the virus through the NPC and into the nucleus. CA-dependent host cofactors include CypA, TNPO3, nucleoporin 358 (Nup358), Nup153, and, most recently, cleavage and polyadenylation specificity factor 6 (CPSF6). The interaction of these cofactors with the virus and their effects upon nuclear entry are discussed in detail in their respective sections.

1.2.9 Integration

1.2.9.1 The mechanism of HIV-1 integration

Once inside the nucleus, HIV-1 DNA can integrate into the host genome. The mechanism of integration was elucidated by analysing the intermediates of *in vitro* integration reactions, which minimally require viral DNA, target DNA, recombinant viral IN, and buffers (126-129). The two blunt ends of the viral dsDNA are bound by a homotetramer of IN, forming a complex called the intasome. Firstly, IN catalyses the removal of the two 3' nucleotides of both DNA ends, creating a 2nt 5' overhang at each end (figure 10A, step 1). IN then catalyses the DNA-strand transfer, in which the free 3' ends attack a pair of phosphodiester bonds of opposite sites of the backbone of the target DNA that are 5nts apart (figure 10A, step 2). This results in covalent bonding of the viral and host DNA flanked by a 5nt duplicated region of host DNA. Host DNA repair enzymes then remove the 2nt 5' overhangs, ligates the single-strand breaks, and fills in the 5nt gaps (figure 10A, step 3). This integrated provirus is a permanent part of the target cell's genome and is copied into any daughter cells during replication (42, 130).

HIV-1 IN is a member of the family of polynucleotidyl transferases and is composed of 3 structurally distinct domains, with the central domain providing the catalytic core. Although several partial structures of IN have been described, the structure of the whole intasome (IN in complex with DNA) remains unsolved. However, the intasome of Prototype Foamy Virus (PFV) has been solved (131, 132) and was used to model HIV-1 IN inhibitors (described below) suggesting they are very similar structures. In the intasome IN forms a homotetramer; two monomers dimerise via their core domains and two of these dimers then dimerise via their NTD and core domains, forming a homotetramer. Of these four subunits, only the inner two are catalytically active, as depicted in figure 10B. The outer two subunits are structurally slightly different, are catalytically inactive, and their precise role is unknown (76). The bound DNA is severely distorted in the intasome such that each active site binds
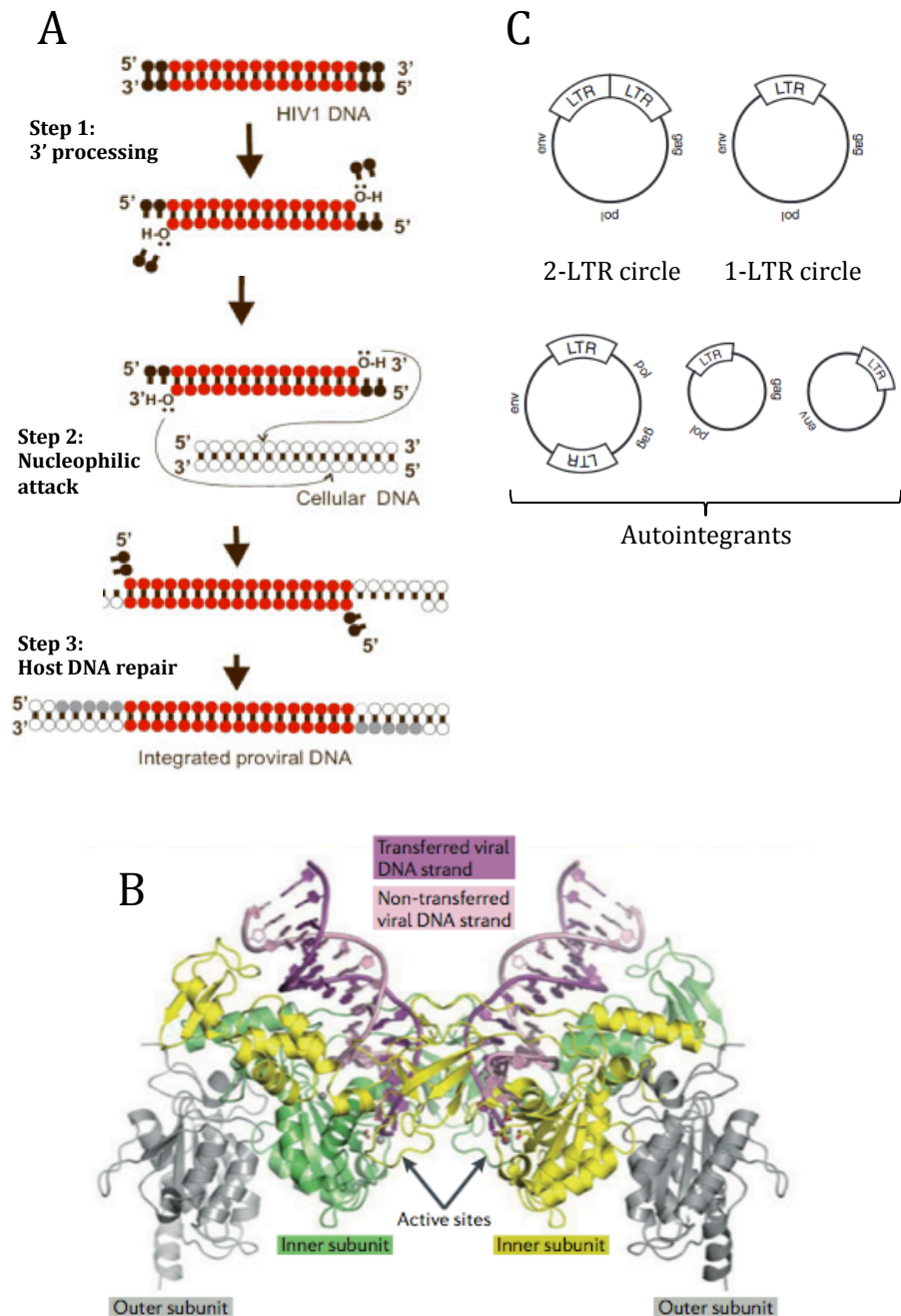
Figure 10. HIV-1 integration
(A) The process of integration of HIV-1 reverse transcripts into host DNA in three steps. Viral DNA shown in red/black, where black is processed DNA. Host DNA in white/grey, where grey is duplicated and repaired by host enzymes. Two black dots represent free electrons in an oxygen atom. Diagram adapted from Savarino 2007. 10B from. **(B)** Crystal structure of the intasome of PFV (prototype foamy virus) (PDB 3OY9, from Hare et al. 2010), which is thought to be very similar to the unsolved HIV-1 intasome. It is an IN heterotetramer (a dimer of dimers) in complex with two DNA molecules. The outer grey subunits are inactive. The inner green/ yellow subunits are catalytically active and form two active sites opposite each other. DNA is shown in purple. Image taken from Engelman and Cherepanov 2012. **(C)** Schematic diagram of the different possible forms of unintegrated HIV-1 DNA. Image adapted from Meyerhans *et al.* 2003.

one end of the viral DNA. A single active site catalyses both the 3' processing and DNA-strand transfer reaction for the end of DNA it binds. The distance between these two active sites dictates the 5nt distance between the two phosphodiester bonds attacked during integration. Raltegravir is the only IN inhibitor (INI) or integration strand transfer inhibitor (InSTI) currently licensed for HIV-1 treatment (133). It binds to the intasome, interacting with the DNA and with the Mg^{2+} ions within the IN active site (77).

Viral DNA that enters the nucleus but does not successfully integrate into the host genome is circularised, forming one of several potential dead-end viral products depicted in figure 10C (134). The non-homologous end joining (NHEJ) DNA repair pathway can ligate viral LTRs to each other, forming 2-LTR circles (135). This is known to only occur within the nucleus and so measurement of the 2-LTR circles is often used as a measure of nuclear entry. Alternatively, 1-LTR circles can form as a result of homologous recombination between the two LTRs or of stalled reverse transcription products that are incapable of integration (40). Thirdly, the 3' ends of viral DNA can attack their own genome, which leads to a variety of circular DNA products known as autointegrants (42). Although these circular forms of DNA cannot integrate, they can persist for several weeks in cells in culture. Circles can even be transcribed but, for unknown reasons, they appear to produce very low levels of Rev, the viral exporter of unspliced mRNA. This means that even if other genes are transcribed they are rarely exported from the nucleus and translated (136) (for details of Rev function, see section 1.2.10.3).

1.2.9.2   The characteristics of HIV-1 integration sites

HIV-1 does not integrate into the human genome at random nor can its pattern of integration be explained by its weak preference for palindromic sequences (137, 138). Instead, there appear to be many different viral and host proteins and genomic features that can influence integration site selection, which have been extensively studied. After the sequence of the human genome was completed in 2001 (139), it became possible to map integration sites onto this

genome and to measure the genomic features surrounding them. Schroder, et al. (140) reported that (in SupT1 cells) HIV-1 preferentially integrates inside genes, in gene-rich regions, and into transcriptionally active regions of the genome. Many subsequent studies have also measured gene density surrounding HIV-1 proviruses and all cell lines and primary cells (including macrophages and activated and resting T cells) tested have shown a strong preference for integration into gene-rich regions (141, 142). Notably, the gene density was similar whether the viruses entered cells via the HIV-1 R5 Env or was pseudotyped with the Vesicular Stomatitis Virus (VSV) 'G' envelope protein (VSV-G), suggesting that the route of entry does not alter integration site selection (142).

Wang et al. (138) completed the largest HIV-1 integration study thus far by examining over 40,000 integration sites of two different HIV-1 strains and confirmed the preference of the virus for integration inside genes and in gene-rich regions. They also identified and/or confirmed that several other genomic factors are associated with HIV-1 integration; HIV-1 integrates into regions with a high density of genomic features such as DNase I hypersensitivity sites (DHSs), CpG islands, and Alu elements (140, 143), but with a low density of features such as LINEs (138). However, each of these genomic features correlates with gene density. For example, hypersensitivity to DNase I indicates that DNA is uncondensed and accessible to DNA binding proteins such as transcription factors and thus DHSs are markers of transcriptionally active DNA (144). CpG islands are also associated with transcription factor binding and are found at 70% of known promoters in the human genome (145). Methylation of cytosines within CpG islands leads to silencing of the associated promoter and so it follows that HIV-1 integration negatively correlates with CpG methylation (138). Alu elements are a type of small interspersed element (SINE) and are the most abundant transposable elements in the human genome. They are sites of transcription factor binding and so it is unsurprising that they are particularly abundant near genes and in transcriptionally active regions (146). In contrast to these SINEs, long interspersed element (LINE) density is negatively correlated with both gene density (147) and with HIV-1 integration frequency

(138). The fact that each of these features correlates with gene density and with each other means that it is extremely difficult to discern which characteristics actually influence the virus.

1.2.9.3 LEDGF/p75 determines HIV-1 integration inside transcription units

Integration inside genes has been attributed to interaction of the virus with the cofactor lens epithelium-derived growth factor p75 (LEDGF/p75), a transcriptional co-activator encoded by the gene *PSIP1* (148). LEDGF/p75 has a chromatin-binding domain in its NTD, followed by an NLS, two A/T-hook motifs, and a C-terminal direct IN binding domain (IBD) (149-151). A crystal structure of the LEDGF/p75 CTD in complex with a dimer of IN core domains has now been solved (152). Knockout of LEDGF/p75 (in murine cells), RNAi-mediated depletion of LEDGF/p75 (in human cells), or expression of LEDGF/p75's IBD (which acts as a dominant negative) causes a post-nuclear entry block to HIV-1 infectivity (153-155). LEDGF/p75 inhibitors, named LEDGINs, bind to IN and inhibit its interaction with LEDGF/p75, thereby potently inhibiting HIV-1 replication (156).

As well as improving the efficiency of integration, LEDGF/p75 also dictates the location of HIV-1 integration. LEDGF/p75 binds to DNA primarily within genes and depletion or knockout of this cofactor prevents HIV-1 from preferentially integrating inside genes (153, 155). LEDGF/p75 is able to interact with both chromatin and IN simultaneously (153) and is required to do this in order to enhance HIV-1 infectivity and integration inside genes (154). These findings lead to the hypothesis that LEDGF/p75 tethers IN to genomic DNA, increasing the efficiency of the integration reaction and directing integration to LEDGF/p75 binding sites. This has been tested by fusing the IBD of LEDGF/p75 to other DNA binding proteins (DBPs) and demonstrating that the integration pattern of HIV-1 follows the binding pattern of that DBP (157). For example, in one study, the chromatin binding domain of LEDGF/p75 was exchanged for heterochromatin protein 1 (HP1). This protein binds sites of H3K9 methylation and is thus a marker of transcriptional repression. When this fusion protein was

overexpressed HIV-1 preferentially integrated outside genes and in regions of low transcriptional activity, confirming the tethering hypothesis (158).

Importantly, although IN-LEDGF/p75 interaction dictates integration into genes, it is not sufficient to ensure integration inside gene-rich areas. CA mutations or depletion of cofactors whose use is dictated by CA can lead to a dramatic reduction in gene density surrounding proviruses without affecting the proportion of proviruses within genes (79, 159). In agreement with this, statistical modeling suggests that LEDGF/p75 affects HIV-1 integration site selection independently of all other factors proposed thus far (138).

The role of LEDGF/p75 in HIV-1 integration may help to explain the virus's preferences with respect to G/C content, which varies depending on the size of the window surrounding a provirus that is examined. HIV-1 preferentially integrates into regions that are G/C-rich across an area of 250Kbp-5Mbp, which is likely because G/C-content positive correlates with gene density. However, the virus favours integration into smaller 50bp-100Kbp A/T-rich regions within this larger G/C-rich region. Local A/T-rich regions of DNA may be targets of integration because they are sites of binding for the A/T-hook motifs within LEDGF/p75 (138).

1.2.9.4 *Genome structure can determine HIV-1 integration site selection*

The structure of DNA within chromatin has been widely proposed to affect integration site selection. Very densely packaged DNA is thought to not be accessible to the virus because extensive regions of heterochromatin such as centromeres and telomeres are cold spots of HIV-1 integration (138, 160). However, within euchromatin, integration is favoured in DNA associated with nucleosomes. This is because DNA is severely bent during the integration reaction and the distortion of DNA on the surface of nucleosomes enhances the efficiency of integration, which has been demonstrated through *in vitro* integration reactions (161, 162). The preference of HIV-1 for A/T-rich DNA may similarly be explained by its increased flexibility (138). As well as the structure

of nucleosomal DNA, histone modifications have been linked to integration site selection. Wang, et al. (138) reported a strong positive correlation between HIV-1 integration sites and markers of transcriptionally active chromatin (such as H3K4 trimethylation, H3K9 and H3K14 acetylation, and H4 acetylation) and a negative correlation between HIV-1 integration sites and markers of transcriptional repression (such as H3K27 trimethylation). These relationships between histone modifications and integration site selection significantly improved the ability of the model designed by Berry, et al. (163) to predict HIV-1 integration patterns (138).

1.2.9.5   Other determinants of HIV-1 integration site selection

Although the major viral determinant of integration site selection is IN (primarily due its interaction with LEDGF/p75), HIV-1/MLV chimeras have been used to demonstrate that CA also significantly influences integration preferences (164). CA mutations have been shown to alter integration site preferences, particularly with respect to gene density (79). This is likely due to the fact that CA dictates interaction with early HIV-1 cofactors, such as CypA, TNPO3, and Nup358, whose depletion also changes the average gene density surrounding HIV-1 integration sites (159) (the effects of these cofactors upon HIV-1 infection are discussed in detail in their respective sections). The most accurate model of HIV-1 integration site selection is still unable to accurately predict integration patterns; there are still several regions of unexpectedly high/low integration frequency (138). There must therefore be other viral and/or host factors influencing integration site selection that have not yet been identified.

1.2.9.6   The consequences of HIV-1 integration site selection

It is often proposed that integration into transcriptionally active regions is beneficial to the virus because it enhances the transcriptional output of the virus. Microarrays of HIV-1 infected and uninfected cells revealed that the median level of transcription is 1.6-3 times higher for viral genes than host genes (140, 165). However, this difference can be up to 75-fold in the absence

of Tat, a viral protein that enhances transcription from HIV-1 LTRs (for details of Tat function see section 1.2.10.2). This dramatic difference was independent of histone modification but was dependent upon the genomic features described previously (as discussed, it is not possible to distinguish between the effects of each of these genomic features because they correlate so strongly with each other) (166). These findings suggest that the local environment of a provirus can dramatically alter its transcription level but that Tat dramatically enhances the transcriptional activity of all proviruses, reducing the influence of the local environment. It remains debatable whether the modest increase in transcription of proviruses in gene-rich regions would have a significant impact *in vivo*. Although it is generally agreed that integration into a transcriptionally active region is beneficial for the virus, there is also evidence that integration into a region that is too highly transcribed can negatively impact the virus through transcription inference (TI) (164). If a provirus is located near to a highly transcribed gene in the same orientation as this gene then transcriptional read-through can occur (167) and if the provirus is in the opposite orientation then convergent transcription can occur (168).

During HIV-1 latency, the provirus is not highly transcribed. It has therefore been hypothesised that the site of integration could determine whether a provirus becomes latent (169). However, there is accumulating evidence against it being a major determinant of latency. Sherrill-Mix et al. (170) recently reported a study of HIV-1 integration sites in 5 different latency models. Two of the models showed a negative correlation between integration inside genes and latency but the other 3 model showed no such trend. Similarly, only one model showed a negative correlation between gene density and latency, with the others showing no correlation. These conflicting results suggest that some (or perhaps all) of the models of latency currently in use are not representative of latency *in vivo*. Studying latency *in vivo* is difficult because cells containing latent proviruses are so rare, but one study of 74 intact proviruses from the resting CD4+ cells of 16 patients showed that 93% lay within actively transcribed genes (141). Although integration into genes and gene density may

be a contributing factor to whether a proviruses becomes latent or not, there are clearly other factors involved.

1.2.10 Transcription and translation of the HIV-1 provirus

1.2.10.1 HIV-1 transcription and splicing

The proviral LTR contains a very strong promoter, from which the host RNA polymerase II (RNAPII) begins transcription. Around 40 different mRNA species are produced from the HIV-1 provirus due to splicing between 4 donor 5' splice sites and 8 acceptor 3' splice sites. These transcripts are categorised into three groups; unspliced ~9Kbp species that encode Gag and Gag-Pol and are also the viral genome, incompletely spliced ~4Kbp species that encode Vif, Vpr, or Env and Vpu, and completely spliced ~1.8Kbp species that encode Tat, Rev, or Nef. All HIV-1 mRNAs are 5' capped and have a 5' untranslated region (UTR) comprising R and U3 regions and a 3' UTR comprising U5 and R regions, which contains 3' processing and polyadenylation signals (171). Host cell mRNAs are almost always completely spliced, and incompletely spliced/unspliced mRNAs are retained in the nucleus due to recognition of splice sites by host proteins, which prevent their export from the nucleus. Therefore, immediately after integration, only the fully spliced mRNAs encoding Tat, Rev, or Nef can be exported and translated. Tat dramatically enhances the efficiency of transcriptional elongation of the HIV-1 provirus and Rev allows export of the other viral mRNAs that still contain splice sites, as described below.

1.2.10.2 Tat and TAR

Tat, the trans-activator of transcription, is not required for the initiation of transcription by RNAPII but it dramatically increases the efficiency of elongation. This means that Tat also encourages the production of more Tat through positive feedback. Tat does this by interacting with a region of the RNA being synthesised known as transactivation responsive region (TAR). TAR is a highly structured RNA sequence that lies just downstream from the transcription start site (+1 to +59). Tat binds to TAR, causes a conformational change in the RNA, and recruits the host positive transcription elongation factor

b (p-TEFb). p-TEFb is a cyclin-dependent kinase (CDK) comprising the catalytic subunit CDK9 and the regulatory subunit cyclin T1 (CycT1) (36). A crystal structure of this complex revealed that, whilst the NTD of Tat makes extensive interactions with CycT1, the central core domain of Tat interacts with RNA (172). In this complex, CDK9 is activated and phosphorylates many host proteins involved in regulating transcription. For example, it phosphorylates the negative elongation factor E (NELF-E), causing it to dissociate from TAR. It also phosphorylates DRB sensitivity-inducing factor (DSIF), converting it from a negative to a positive elongation factor. It also phosphorylates the CTD of RNAPII-A, the initiating form of RNAPII, and converts it into RNAPII-0, the elongating form (36). Through such actions the Tat-p-TEFb complex greatly enhances the efficiency of elongation during transcription of the HIV-1 provirus.

1.2.10.3 Rev and the RRE

Normally, incompletely spliced or unspliced mRNAs are retained in the nucleus and eventually degraded as a part of normal cell turnover. Rev prevents this from happening to HIV-1 mRNAs by binding to the Rev responsive element (RRE), a sequence that lies within an intron in *env* and is present in all incompletely spliced or unspliced viral mRNAs. Like TAR, the RRE forms a stable stem-loop structure in RNA and, like Tat, Rev binding to the RRE causes conformational changes in the RNA structure (36). Binding of Rev to the RRE RNA causes Rev to dimerise and form a V-shaped structure, in which the two monomers interact with adjacent sites in the RRE RNA. This dimerisation creates a new protein-protein interface, which causes other Rev dimers to bind cooperatively and thus form higher order Rev oligomers along the RRE. Figure 11 shows how Rev oligomers simultaneously interact with RNA and Crm1, the exportin that mediates nuclear export of the bound viral mRNA (173).

1.2.10.4 HIV-1 translation and frameshifting

All viral mRNAs are translated on cytosolic polysomes except for Env and Vpu, which are encoded on the same mRNA and are translated at the rough ER (43). The longest mRNA encodes Gag and Gag-Pol. One in 20 times that this mRNA is
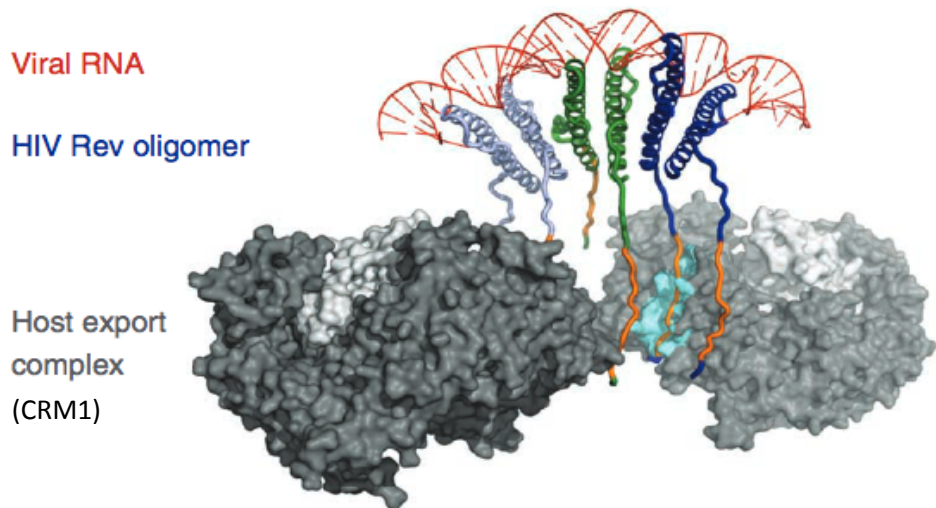
Figure 11. Rev and the RRE
Model of a Rev hexamer bound to two RNA molecules, which represent part of the RRE, and two Crm1 molecules. Rev dimers shown in lilac, green, or blue, with their Crm1-binding NESs shown in orange. RNA is shown in red. Crm1 monomers shown in grey with their Rev-binding domain in cyan. Image taken from Daugherty et al. 2010.

translated, a +1 frameshift occurs at the end of Gag, causing readthrough into Pol and synthesis of the Gag-Pol polyprotein. This frameshift is due to two elements, a slippery U-rich sequence where the actual framshift occurs and a pseuodknot structure in the RNA that causes the RNAPII to pause at the slippery sequence thus increasing the chances of slipping (36).

1.2.11 HIV-1 virion assembly, budding, and maturation

1.2.11.1 Assembly at the plasma membrane

As an enveloped virus, HIV-1 assembles at the plasma membrane of infected cells and buds from its surface. Gag and Gag-Pol are polypeptides during assembly and are not cleaved into their constituent parts until maturation, which occurs during and/or after budding. Whereas cytoplasmic Gag is soluble and monomeric (or in low order multimers), it extends into a rod-shape structure at the plasma membrane (174, 175). Gag is thought to be targeted to the plasma membrane by its MA domain. tRNAs interact with the N-terminus of MA and prevent it from interacting with membranes that lack the lipid raft component $PI(4,5)P_2$ (176). Interaction of MA with $PI(4,5)P_2$ induces a conformational change known as the 'myristoyl switch' because it causes the N-terminal myristoyl group of MA to be exposed and to anchor Gag stably in the membrane (177, 178).

Live-cell imaging has shown that small numbers of Gag molecules interact with HIV-1 genomic RNA in the cytoplasm and move to the plasma membrane together before recruiting many more Gag molecules from the cytoplasmic pool (179). The HIV-1 genome dimerises in the cytoplasm via a sequence in its 5' UTR called the dimer initiation signal (DIS). Several genetic studies have suggested that the NC domain of Gag interacts with a highly structured packaging signal (ψ), which is only present in the viral genome (and not any other viral RNAs), thereby ensuring that only full-length HIV-1 transcripts are incorporated into particles (43). Kutluay et al. (176) have recently shown that whilst this is true for cytoplasmic Gag, multimeric plasma membrane-bound Gag interacts with viral gRNA much more extensively and that ψ becomes less

important. This finding suggests that conformational changes occur in Gag during virion assembly that encourage it to interact with the viral genome.

Env reaches the plasma membrane independently of Gag. Having been co-translationally inserted into the membrane of the rough ER, Env follows the secretory pathway to the plasma membrane (43). Here, the MA domain of Gag recruits Env by interacting with the cytoplasmic tail of Env and is thought to be essential for incorporation of 10-15 Env trimers into the budding virion (180).

As in mature capsids, the CA CTD mediates Gag-Gag interactions in assembling immature virions. Due to these interactions Gag forms a hexameric lattice with the molecules orientated radially (MA on the outer layer at the plasma membrane and p6 on the inner layer) as show in figure 12A. Whereas the mature lattice bends due to pentamers, the immature core bends due to large gaps/holes (181) (figure 12B). The innermost domain of Gag is p6, which is thought to be required for incorporation of Vpr (182) and is involved in recruiting endosomal sorting complexes required for transport (ESCRT) machinery, as described below.

1.2.11.2 HIV-1 and the ESCRT pathway

The ESCRT pathway catalyses fission of membrane necks, for example during the release of vesicles inside multivesicular bodies (MVBs) and during separation of daughter cells during cell division (183). The ESCRT pathway involves over 30 proteins that form the key complexes; ESCRT-0, I, and II complexes interact with scaffold proteins and orchestrate formation of downstream complexes, ESCRT-III complexes interact with membranes and form helical/spiral assemblies that constrict the membrane neck and induce fission, and the ATPase vacuolar protein sorting 4 (VPS4) is required to provide the energy for remodeling and recycling of ESCRT-III proteins (184). HIV-1 exploits this complex system, recruiting components of the pathway to Gag during assembly. The p6 domain of Gag has two late (L) domains that recruit the ESCRT-I component tumour susceptibility gene 101 (TSG101) and ALG-2-
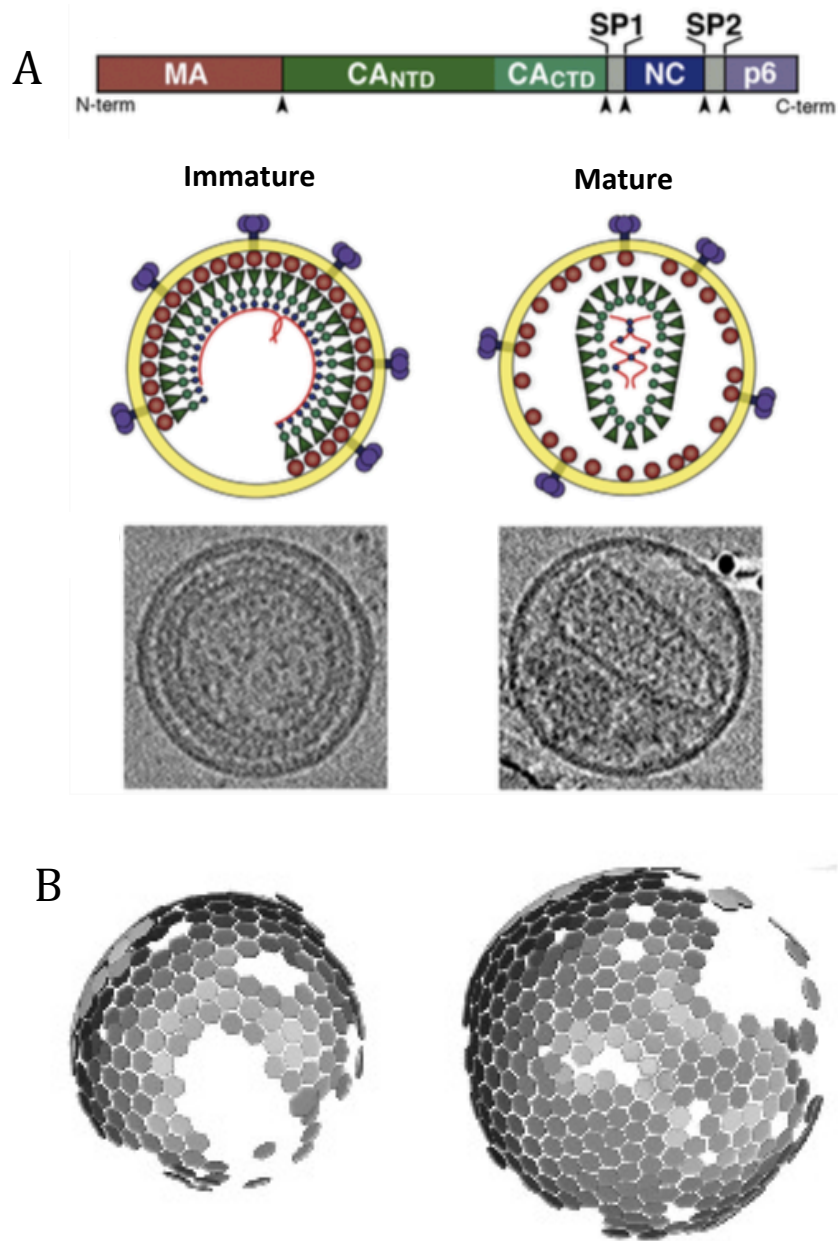
Figure 12. HIV-1 maturation
(A) (Top) Schematic diagram of the domains of uncleaved Gag polyprotein. Arrows indicate sites of PR-mediated cleavage during maturation. (Middle) Cross-sectional diagrams of immature or mature virions, with colours corresponding to the above schematic. DNA is depicted as red lines, IN as blue circles, the membrane in yellow, and Env as purple circles. (Bottom) EM cross-sectional images of immature or mature virions. Image taken from Ganser-Pornillos et al. 2008. **(B)** Model of immature viral cores based upon subtomogram averaging, revealing that they are incomplete with many holes/defects. Image taken from Briggs et al. 2009.

interacting protein X (ALIX) (185-187), which in turn recruit the VPS4 (188, 189) and the ESCRT-III components chromatin modifying proteins 2A and 4B (CHMP2A and CHMP4B) (190). Figure 13 shows how the ESCRT machinery is thought to assemble on the inside of the budding virion, with CHMP2A/4B forming the spiral inside the membrane neck and VPS4 providing the energy for depolymerisation of this spiral and thus constriction of the neck, which leads to membrane fission (191).

1.2.11.3 HIV-1 virion maturation

The immature virion must undergo a series of dramatic conformational changes in a process known as maturation in order to become infectious. This begins during or immediately after budding and requires activation of the viral PR. In the context of Gag-Pol, PR is monomeric and dimerisation is required to form the active site of this aspartic acid protease. Transient dimerisation of PR in the context of Gag-Pol leads to autocleavage of PR from Pol. The PR dimer is now stabilised and can catalyse cleavage of the other Gag-Pol cleavage sites, which it recognises by their secondary structure rather than their sequence. PR cleaves 5 sites in Gag leading to the production of MA, CA, NC, and p6, as well as release of the spacer peptides SP1 and SP2. These peptides are not thought to have any role in HIV-1 infection other than to assist the correct folding of Gag proteins during this cleavage process. Gag-Pol is cleaved in 5 sites, leading to the production of p6, PR, RTp66, and IN, as well as the release of the short transframe protein (TFP), which is a result of the frameshift occurring between Gag and Pol (43). The cleavage sites are processed with very different efficiencies, which dictate the order of cleavage. SP1/NC cleavage is required to activate Env, although the mechanism of this is not fully understood (192), MA/CA cleavage disassembles the immature lattice leaving a membrane-bound MA layer, CA/SP1 cleavage produces free CA that can form the mature conical core, and NC/SP2 cleavage is required for release of NC and formation of the NC/RNA ribonucleoprotein complex (43).
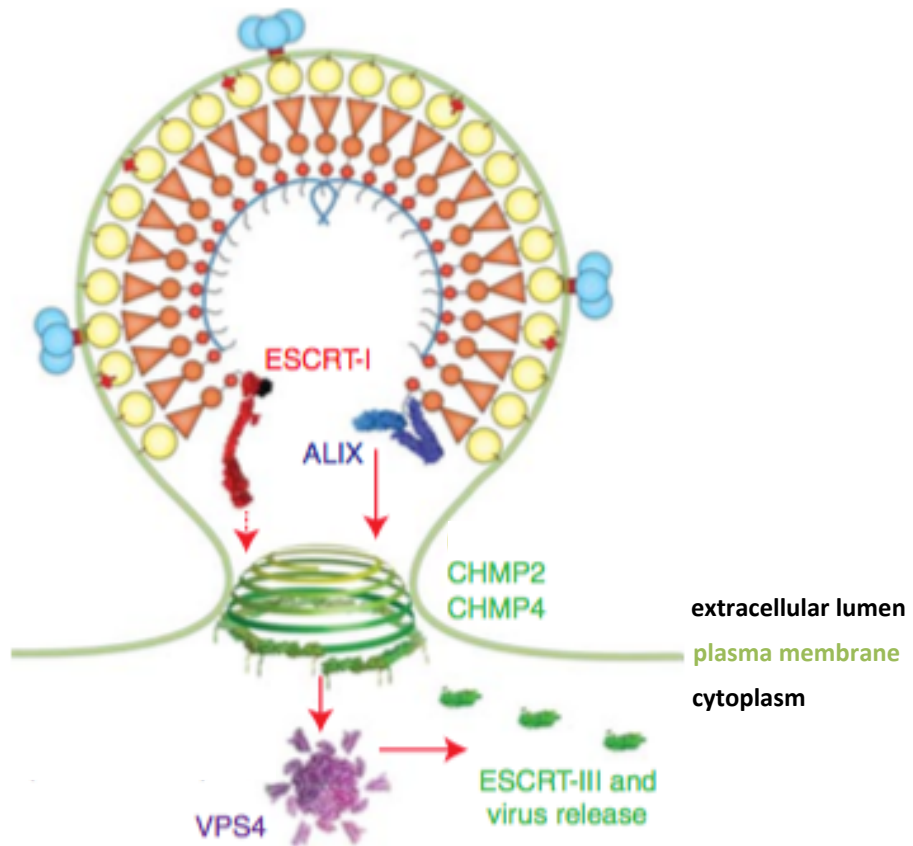
Figure 13. HIV-1 budding and the ESCRT pathway
Simplified model of an assembled virion budding at the plasma membrane through hijacking host ESCRT machinery. Gag comprises yellow circles of MA, orange triangles of CA, orange circles of NC, and red circles of p6. Motifs in p6 recruit the TSG101 subunit of the heterotetrameric complex ESCRT-I and the V domain of ALIX. These interactions lead to recruitment of ESCRT-III proteins CHMP1, 2, and 4, which assemble into a dome-shaped structure at the neck of the budding virion. This recruits VPS4, an ATPase, which is required for depolymerisation of the dome. This causes the neck to constrict and the plasma membranes to fuse. Image taken from Sundquist and Krausslich 2012.

Maturation can be inhibited using two classes of ARVs, protease inhibitors (PIs) and the newly developed maturation inhibitors. There are 10 PIs currently licensed as ARVs, all of which interact with the active site of the PR dimer and directly occlude substrate binding. Resistance mutations can of course arise within this active site but these often lead to significant fitness defects. Compensatory mutations can arise within the active site and/or within cleavage sites (77). Recently, the maturation inhibitor Bevirimat was discovered to inhibit HIV-1 by binding to the CA-SP1 site rather than to PR. Bevirimat, which is not licensed as an ARV drug, prevents cleavage of this site and thus leads to stabilisation of immature uninfectious capsids (193).

1.3 HIV-1 restriction factors

As described above, HIV-1 has evolved many interactions with host proteins and to exploit many host processes that are advantageous to the virus. Conversely, host species have evolved proteins that interact with the virus or disrupt processes essential for virus replication. In the case of HIV-1, several of these restriction factors have been extensively characterised, although there are certainly others that remain unidentified. Many of the restriction factors described below not only inhibit the virus but also act as pattern recognition receptors (PRRs) and trigger innate immune signaling, which induces an antiviral state.

1.3.1 TRIM5α and TRIMCyp

Stremlau *et al.* (104) identified TRIM5α as a primate species-specific restriction factor of HIV-1. Tripartite motif (TRIM) proteins comprise a really interesting new gene (RING) domain with E3 ubiquitin ligase activity, a coiled-coil domain that mediates multimerisation, and a type 2 B-box zinc finger domain. TRIM5α consists of an N-terminal TRIM domain and a C-terminal PRYSPRY domain. This PRYSPRY region interacts directly with HIV-1 CA and determines the viral specificity of the restriction factor, so it is under strong positive selection pressure (194). Two separate retrotransposition events in different species of monkey have lead to the fusion of the TRIM domain of TRIM5α to the HIV-1 cofactor CypA (which can interact directly with CA, as detailed in section 1.5.2.2) and thus the creation of the restriction factor TRIMCyp (195, 196).

TRIM5α inhibits HIV-1 infection prior to reverse transcription. It has been proposed that TRIM5α forms a hexagonal cage around the viral core (197) as shown in figure 14A and induces core disassembly (114). Interestingly, in the presence of proteasome inhibitors or RING domain E3 ligase mutations, TRIM5α allows reverse transcription to proceed and does not affect core disassembly, but still inhibits the virus prior to nuclear entry (198, 199). This has lead to the hypothesis of a two-step mechanism in which simply binding to the core is sufficient to inhibit infection, but another proteasome-dependent
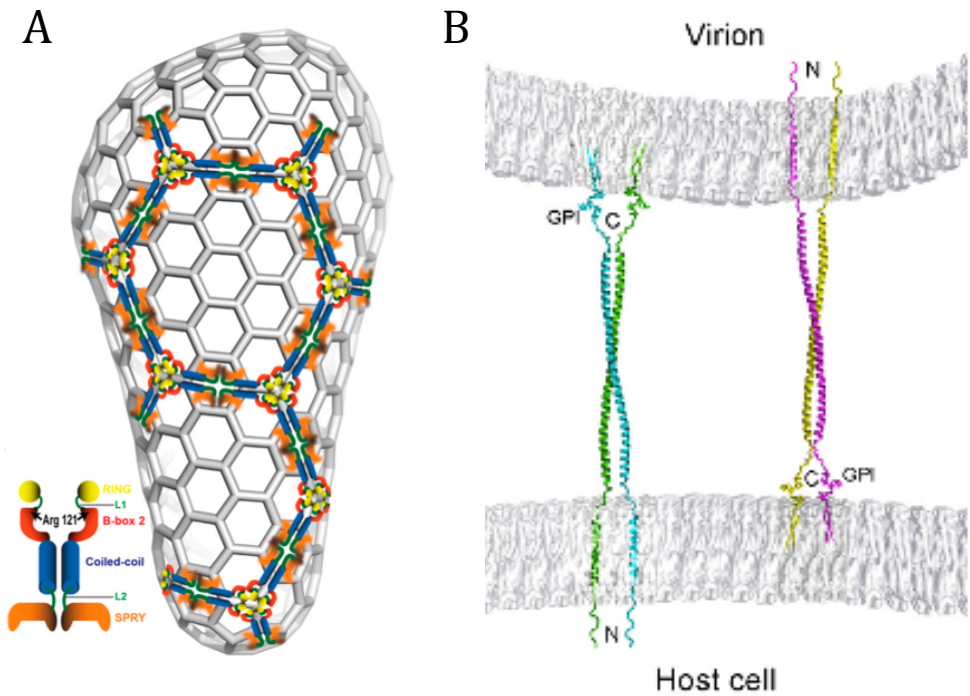
Figure 14. Restriction factors of HIV-1
(A) Model of a hexameric lattice/cage of rhTRIM5α dimers around the mature HIV-1 core. Image taken from Ganser-Pornillos 2011. **(B)** Model of two tetherin homodimers in different orientations, both of which are attaching virion and host membranes together. Lipid bilayers are shown in grey. C and N termini are labeled as C and N, respectively. GPI = glycophosphatidylinositol anchor. Image (using PDB 3MQC) taken from Yang et al. 2010.

step is required to induce core disassembly and to inhibit reverse transcription (200, 201). It has also recently been shown that TRIM5α induces innate immune signaling via activator protein 1 (AP-1) and nuclear factor κB (NFκB) by producing unlinked K63 ubiquitin chains that are recognised by tumour growth factor-β activated kinase 1 (TAK1). Recognition of CA cores enhances this signaling, suggesting that this restriction factor is also a PRR (202, 203).

1.3.2 Tetherin

A virion-tethering restriction factor of HIV-1 was sought after the observation that deletion of the accessory protein Vpu prevented efficient release of budding virions from the surface of interferon (IFN) stimulated cells (204-206). Using microarrays to identify candidate genes, two groups simultaneously identified the IFN-inducible transmembrane protein tetherin, also known as bone marrow stromal cell antigen 2 (BST-2), as a novel HIV-1 restriction factor (207, 208). IF and EM have since been used to show the accumulation of viral particles at the cell surface when tetherin is overexpressed or Vpu deleted (208, 209). The NTD of tetherin is an intracellular tail followed by a transmembrane domain. Almost the entire ectodomain is a single long α-helix, which mediates tetherin dimerisation by forming a coiled-coil. The CTD has a small glycophosphatidylinositol (GPI) anchor, which is able to interact with another membrane (210). When tethering a virion to a host cell plasma membrane, tetherin can be found oriented in either direction (as shown in figure 14B) (209). Like TRIM5α, tetherin has been shown to induce innate immune signaling via NFκB after recognition of susceptible HIV-1 mutants (211).

Vpu is a small transmembrane accessory protein that antagonises tetherin. There is evidence that Vpu interacts with tetherin, causes endocytosis of tetherin from the surface, and also that it prevents trafficking of newly synthesised tetherin to the surface, so its exact mechanism of action remains debated (207). Whilst the pandemic HIV-1 group M uses Vpu to antagonise tetherin, non-pandemic HIV-1 group O strains use Nef to do this (212) and HIV-2 uses Env (213). Non-epidemic strains of HIV-1 O and N appear to be unable to

antagonise tetherin as efficiently and this has been suggested as a significant barrier to transmission of these viruses (214).

1.3.3 APOBEC3G

Like tetherin, the restriction factor apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3G (APOBEC3G) was discovered due to the finding that Vif was required for HIV-1 infectivity in a cell-type dependent manner (215). APOBEC3G is an polynucleotide cytidine demainase, meaning it catalyses the conversion of cytidine residues within RNA into uracil, and deoxycytidine residues within single-stranded DNA into an unnatural base that base pairs with adenosine. During reverse transcription, APOBEC3G causes hypermutation of the newly synthesised minus-strand DNA and the resulting unnatural base in the minus-strand causes guanosine to adenosine mutations in the plus-strand DNA (216, 217). Up to 10% of cytidines can be edited during one replication cycle, which leads to genetic instability, but APOBEC3G can also prevent progression of reverse transcriptase (218). APOBEC3G dimers are packaged into HIV-1 virions through interaction with the genomic RNA and with NC. APOBEC3F and H have also been shown to restrict HIV-1 but do so much less efficiently than APOBEC3G and with slightly different local sequence preferences (219). Vif antagonises APOBEC3G and F by recruiting it to the Cullin 5 E3 ubiquitin ligase complex, which results in the ubiquitination and proteasomal degradation of protein (220).

1.3.4 SAMHD1

Some myeloid cells, such as monocytes and dendritic cells, are very poorly permissive to HIV-1 infection, blocking infection post-entry and pre-reverse transcription. This restriction was known to be antagonised by viral protein x (Vpx), an accessory protein encoded by HIV-2 and some SIV strains (221). Two separate studies of Vpx-binding partners simultaneously identified sterile α motif and HD domain-containing protein 1 (SAMHD1) as the myeloid-specific restriction factor (222, 223) and it has since also been found to restrict infection in resting CD4+ T cells as well (224). SAMHD1 is thought to inhibit

HIV-1 reverse transcription non-specifically, by reducing intracellular dNTP levels using the deoxynucleoside triphosphate triphosphohydrolase activity of its HD domain (225). Interestingly, overexpression of SAMHD1 does not confer resistance to HIV-1 in actively dividing cells (222, 226) and so regulation of this restriction factor is still under active investigation.

The NTD of Vpx directly interacts with the CTD of SAMHD1 and recruits the E3 ubiquitin ligase DCAF1, leading to the ubiquitination and proteasomal degradation of SAMHD1 (227). HIV-1 does not encode Vpx nor any other protein that can counteract SAMHD1. Due to the fact that infection with HIV-2, which does encode Vpx, leads to lower viral loads than HIV-1, it has been hypothesised that sensitivity to SAMHD1 may be beneficial to the virus (228). This is supported by the finding that, in the presence of exogenous Vpx, HIV-1 triggers a type I IFN response in myeloid dendritic cells (DCs) (93). By avoiding infection of such myeloid cells, HIV-1 may reduce its recognition by the immune system and so actually spread more efficiently.

1.4 HIV-1 accessory proteins

As well as encoding structural proteins in Gag, enzymes in Pol, the viral envelope proteins in Env, and the regulatory proteins Tat and Rev, HIV-1 encodes the accessory proteins Nef, Vif, Vpr, and Vpu. These proteins are not essential for virus replication *in vitro*, but nevertheless play very important roles in HIV-1 infection *in vivo*. As described above in sections 1.3.3 and 1.3.2, Vif and Vpu antagonise the host restriction factors APOBEC3G and tetherin, respectively. Additional functions of these accessory proteins are discussed below.

1.4.1 Nef

Nef (negative regulatory factor) is a small non-enzymatic accessory protein that is not essential for HIV-1 replication but significantly increases viral infectivity *in vitro* and enhances pathogenicity *in vivo* (229). A small number of patients have been founded to be infected with viruses containing large deletions in Nef, all of which are either elite controllers, long-term non-progressors (LTNPs), or slow progressors (SPs) (230). Nef purportedly plays multiple different roles in HIV-1 infection and although they can be separated genetically, it has not been possible to determine which of these functions allow control of Nef-deficient viruses *in vivo.* It has recently been shown that infectivity enhancement of Nef is dependent upon dynamin 2 and clatherin (231). Accordingly, Nef has previously been implicated in the clatherin-mediated endocytosis of both CD4 and major histocompatibility complex (MHC) class I molecules from the surface of infected cells. Nef induces endocytosis of CD4 by interacting directly with the cytoplasmic tail of CD4 and with activating protein 2 (AP-2), a clathrin adaptor protein. Once endocytosed, Nef also targets CD4 to MVBs for degradation in lysosomes (232). Downregulation of Nef prevents superinfection and so promotes spread of the virus to uninfected cells. Nef also induces endocytosis of MHC class I molecules by interacting with its cytoplasmic domain and also the clathrin adaptor AP-1 (233). This reduces the ability of cytotoxic T lymphocytes (CTLs) to recognise infected cells, which may explain why it is beneficial for Nef to be synthesised so early after integration

(discussed in further detail in section 4.1.2). Due to the fact that CTL responses aren't relevant *in vitro*, it cannot be the ability of Nef to downregulate MHC that causes it to increase infectivity in culture. It is there either downregulation of CD4 or another unidentified factor that causes Nef to increase infectivity of the virus *in vitro*. Nef has also been shown to interact with p21-activated kinase 2 (PAK2), which has been implicated in cytoskeletal scaffolding, and the myeloid specific haemopoietic cell kinase (Hck), which activates the signal transducer and activator of transcription 3 (STAT3) (229). These observations have lead to the controversial hypothesis that Nef induces cellular activation.

1.4.2 Vpu

As described in section 1.3.2, Vpu efficiently anatagonises the host restriction factor tetherin. Like Nef, Vpu is also involved in downregulation of CD4. However, whereas Nef causes endocytosis and degradation of CD4 from the cell surface, Vpu recruits an E3 ligase complex to CD4 in the endoplasmic reticulum (ER), which polyubiquitinates its cytoplasmic tail. This causes newly syntheised CD4 to be retained in the ER and then directs it into the ER-associated degradation (ERAD) pathway (234).

1.4.3 Vpr

Vpr (viral protein r) has been suggested to have multiple roles in HIV-1 infection. The most widely studied effect of Vpr has been its ability to arrest the cell cycle during G_2/M transition (235). It has recently been shown that Vpr induces cell-cycle arrest by interacting directly with SLX4 and activating the SLX4 complex, which is involved in resolving Holliday junctions during homologous DNA recombination. However, Vpr-SLX4 interaction has also been shown to suppress innate immune responses against the virus (236). Furthermore, the ability of Vpr to arrest the cell-cycle has been shown to correlate with an ability to alter nuclear envelope architecture (237), induce apoptosis in infected and bystander cells (238), and to regulate transcription through interaction with the transcriptional co-activator p300 (239). It has therefore been difficult to discern whether any of these phenotypes are

dispensable for viral replication. Vpr has also been implicated in the fidelity of reverse transcription and the nuclear import of viral PICs (240), although these latter observations are highly controversial.

1.5 Cofactors of early HIV-1 infection

Several events in early HIV-1 infection have been extensively studied including the mechanisms of entry, reverse transcription, and integration. However, it remains unclear what happens between these discrete events. Recently, the field has become particularly interested in the composition of the virion throughout early infection, the timing and location of reverse transcription and uncoating, the mechanism by which the virus traverses the NPC, and how the virus selects its site of integration. Several host proteins have been reported to interact with the virus during early infection and to enhance infectivity, as discussed below.

1.5.1 Cofactors of HIV-1 nuclear entry

After reverse transcription and uncoating, the viral PIC enters the nucleus where it integrates into host chromatin. HIV-1 infects non-dividing cells as efficiently as dividing cells and so must be able to enter intact nuclei via the NPC. As discussed in section 1.2.8.1, the NPC is selectively permeable and whilst molecules below ~40kDa can passively diffuse through, larger proteins/complexes must be actively transported through the nuclear pore across a RanGTP gradient. Most cargo that is transported into the nucleus does so in complex with a karyopherin. Karyopherin/cargo complexes are thought to pass through the NPC by making sequential weak interactions with FG repeats from many different nucleoporins (119). It is therefore unsurprising that several karyopherins and nucleoporins have been identified as HIV-1 nuclear import cofactors, details of which are discussed below.

1.5.1.1 TNPO3

TNPO3 (transportin 3, TRN-SR2) is a member of the importin-β family of karyopherins, which import SR (serine/arginine-rich) proteins via their phosphorylated RS (arginine/serine-rich) domains (241). A role for TNPO3 in HIV-1 infection was first reported by Christ, et al. (242), who demonstrated that depletion of TNPO3 caused a substantial reduction in HIV-1 single-round infectivity, and this observation has since been confirmed by two independent

siRNA screens (243, 244). Controversy as to whether TNPO3 depletion inhibits HIV-1 infection prior to nuclear entry (242) or after nuclear entry (245, 246) has been explained by the fact that some quantitative real-time PCR (qPCR) assays for 2-LTR-circles were also detecting autointegrants, which can contain two LTRs but form in the cytoplasm and so do not indicate nuclear entry (discussed in section 1.2.9.1). The development of an assay that exclusively measures 2-LTR-circles confirmed that TNPO3 depletion does indeed inhibit HIV-1 prior to nuclear entry, as expected (247). One study reported that TNPO3 depletion caused CA to accumulate in the nucleus as measured by fractionation and so suggested that TNPO3 was required for CA export (248). However, given prior reports of cores 'docking' at NPCs (discussed in section 1.2.7.1), we interpret this to mean that cores were bound to the outside of the nuclear pore/envelope rather than inside the nucleus. This observation therefore supports the notion that TNPO3 is required for HIV-1 nuclear entry.

As well as preventing nuclear entry, depletion of TNPO3 has been shown to change the integration site preferences of HIV-1. Ocwieja et al. (159) showed that WT HIV-1 integrated into regions with an average of ~21genes/Mb in control HEK-293T cells but only ~11genes/Mb in cells depleted of TNPO3. As expected, this reduction in gene density correlated with a reduction in gene-associated genomic features, such as DNase sites, CpG islands, and GC content. Similar results were observed in HeLa cells (79). Notably, the average gene density surrounding proviruses in TNPO3-depleted cells was still above average for the human genome (~7genes/Mb).

TNPO3 has been shown to interact directly with HIV-1 IN by yeast-2-hybrid assay (242), co-immunoprecipitation (249, 250), and AlphaScreen (amplified luminescent proximity homogenous assay screen) (251). Furthermore, this interaction can be disrupted by binding of RanGTP to TNPO3, consistent with the behavior of nuclear import cargoes (252). However, some mutations that prevent IN-TNPO3 interaction do not inhibit infectivity and so it is thought that TNPO3 may be redundant for the nuclear import of IN (253). In agreement with this observation, the requirement for TNPO3 in HIV-1 infection maps

genetically to CA, not IN (250). TNPO3 has recently been shown to interact directly with CA as well as IN (87, 245) and purified recombinant TNPO3 reportedly stimulates core uncoating *in vitro* (87). A model that takes into account all of these findings is that TNPO3 first interacts with CA and induces uncoating, which exposes IN, which then interacts with TNPO3 or an alternative karyopherin, importing the HIV-1 PIC into the nucleus.

1.5.1.2 Nup358

Nup358 (nucleoporin 358, RAN-BP2) is the largest known nucleoporin and the primary component of the filaments/fibrils that extend from the cytoplasmic face of the NPC into the cytoplasm (119) (as shown in figure 9A). The tips of these filaments interact with microtubules and are hypothesised to coordinate trafficking from microtubules through the NPC (120). Nup358 has four Ran binding domains and plays an important role in the formation and disruption of cargo/karyopherin/Ran complexes. It also comprises a zinc finger domain, an E3 ligase domin, and a cyclophilin domain (254).

Nup358 was identified as an HIV-1 cofactor in two independent siRNA screens (243, 244). It has been shown to be essential for nuclear import of Rev (255, 256), an HIV-1 accessory protein required for Crm1-dependent export of unspliced mRNAs (Rev is discussed further in section 1.2.10.3). Nup358 has also been implicated in the early stages of HIV-1 infection, with Nup358 depletion exhibiting similar phenotypes to TNPO3 depletion. For example, Nup358 depletion impairs nuclear entry as measured by 2-LTR circles (79, 257-259) and leads to the accumulation of CA at the nuclear periphery as measured by IF. This latter effect was shown to be specific to Nup358 and did not occur when depleting other nucleoporins such as nucleoporins 214, 98, or 153 (Nup214, 98, or 153) (258). Furthermore, Nup358 depletion causes similar changes to HIV-1 integration site selection as TNPO3 depletion (79, 159). These similarities between TNPO3 and Nup358 suggest that they both function at the same stage of HIV-1 infection.

The cyclophilin domain of Nup358 (Nup358$_{Cyp}$) interacts directly with the CypA-binding loop of CA and catalyses *cis-trans* isomerisation of P90 (79, 88). However, the significance of this in HIV-1 infection remains unknown. Residue V61 of Nup358$_{Cyp}$ was found to be essential for its interaction with CA and also to be under strong positive selection pressure, which supports the hypothesis that Nup358 acts a cofactor of viral infection through its interaction with CA (79). It has recently been shown that Nup358ΔCyp (a truncated version of Nup358 that has no cyclophilin domain) is sufficient for HIV-1 infection of murine cells (260) and can interact directly with CA-NC *in vitro,* albeit more weakly than the full-length protein (258). This suggests that CA may interact with other regions of this enormous protein, as well as its cyclophilin domain.

1.5.1.3 Nup153

Like Nup358, the nucleoporin Nup153 has also been implicated in HIV-1 infection. There are thought to be several different populations of Nup153 within the nucleus that have different locations and dynamics, and play different roles. It was first identified as a structural component of the nuclear basket (as shown in figure 9A). In the absence of Nup153, NPCs are mobile within the nuclear envelope suggesting that this nucleoporin is needed to anchor NPCs within the membrane (261). However, there is also evidence that Nup153 exists throughout the nucleoplasm and is mobile. Recent chromatin immunoprecipitation (ChIP) analyses revealed that it interacts with up to 25% of chromatin and preferentially associates with actively transcribed regions of the genome (262). The dynamics of Nup153 have been studied using fluorescence recovery after photobleaching (FRAP) and it was found that Nup153 signal returned rapidly (<120secs) after photobleaching. This suggests that Nup153 is highly mobile in comparison to other nucleoporins tested (263, 264). Intriguingly, inhibitors of transcription such as actinomycin D and DRB prevented Nup153 signal from recovering after photobleaching, indicating that Nup153 mobility is dependent upon active transcription. Furthermore, Nup153$_{NTD}$ was found to bind RNA and Nup153$_{CTD}$ to interact with lamin B, a protein involved in chromatin organisation and transcriptional regulation

(263). These observations suggest that Nup153 exists in at least two different populations within the nucleus.

Given these observations and the fact that Nup153 is such a large protein, it is unsurprising that visualisation of Nup153 using immunogold labeling or immunofluorescence produces different results depending upon the antibody-binding site. What is surprising, however, is that antibodies targeting the FG region of Nup153 are found on the cytoplasmic face of the NPC (265). This suggests that the C-terminal FG repeat region of Nup153 is flexible enough to extend through the NPC and make contacts with cytoplasmic factors.

Depletion of Nup153 has been shown to specifically inhibit HIV-1 prior to nuclear entry in several studies (79, 243, 244, 258, 259, 266). As with TNPO3, there has been disagreement as to whether Nup153 interacts with HIV-1 IN (267) or CA (83, 84) *in vitro*. The possibility that both viral proteins interact with Nup153 is supported by the finding that HIV-1 chimeric with either MLV CA or IN is less sensitive to Nup153 depletion than HIV-1, but that HIV-1 must be chimeric with both MLV CA and IN in order to become completely insensitive to Nup153 depletion (266). As with TNPO3, it is possible that Nup153 interacts with both proteins sequentially. Despite the fact that Nup153 has been implicated in the same stage of infection as Nup358 and TNPO3, its depletion does not appear to have a profound effect upon integration site targeting. Whilst one study found that Nup153 depletion caused a subtle but statistically significant reduction in gene density surrounding proviruses (268) another study found no significant difference (84).

1.5.1.4 Other nucleoporins

TNPO3, Nup358, and Nup153 are by far the most well studied nuclear import factors implicated in early HIV-1 infection, but others have been reported. For example, depletion of Nup62 has been reported to specifically inhibit HIV-1 at a post-nuclear entry but pre-integration step. This channel nucleoporin has been shown to interact directly with chromatin and HIV-1 IN simultaneously,

suggesting it may tether the virus to host DNA (269). Nup98 and Nup214, which are components of cytoplasmic filaments, have also been implicated in a post-nuclear entry pre-integration step of HIV-1 infection. However, their depletion inhibits both HIV-1 and MLV (84, 258), suggesting they may aid HIV-1 infection indirectly. Depletion of importin 7 has also been shown to modestly reduce the nuclear import of HIV-1 (270, 271), but the specificity of this has been questioned because it has also been shown to reduce transfection efficiency (272).

1.5.2 Cyclophilin A is a cofactor of early HIV-1 infection

1.5.2.1 Cyclophilin A is a PPIase

The cyclophilins are a family of peptidyl-prolyl isomerases (PPIases), meaning they catalyse conversion between *cis* and *trans* isomers of proline (as shown in figure 15A) (273). They are ubiquitous in prokaryotes and eukaryotes and 17 have been identified in humans so far (274). Each member has a conserved Cyp domain of approximately 109 amino acids, which forms a variation of an 8-stranded antiparallel β-barrel with 2 α-helices closing the barrel on each end (273, 274). Whilst the majority of cyclophilins have other functional domains, four of them, including Cyclophilin A (CypA), are single domain PPIases (274). Most cyclophilins are thought to function as chaperones or foldases (274) and have been implicated in a range of cellular functions including trafficking and cell signaling (275). CypA is the archetypal cyclophilin and the most abundant in the family, constituting up to 0.6% of total cellular protein (275). It is an 18kDa cytoplasmic protein that was first identified as a receptor for the immunosuppressive drug CsA (276) and subsequently as a PPIase (277, 278).

CsA acts as an immunosuppressant by dampening T cell activation. During T cell activation, Ca^{2+} and calmodulin activate calcineurin, a Ser/Thr-protein phosphatase. Calcineurin dephosphorylates the cytoplasmic nuclear factor of activated T-cells (NFAT), which causes a conformational change in NFAT, exposing its NLS and translocating it to the nucleus (279). In the nucleus, NFAT forms a complex with transcription factors belonging to the families FBJ murine
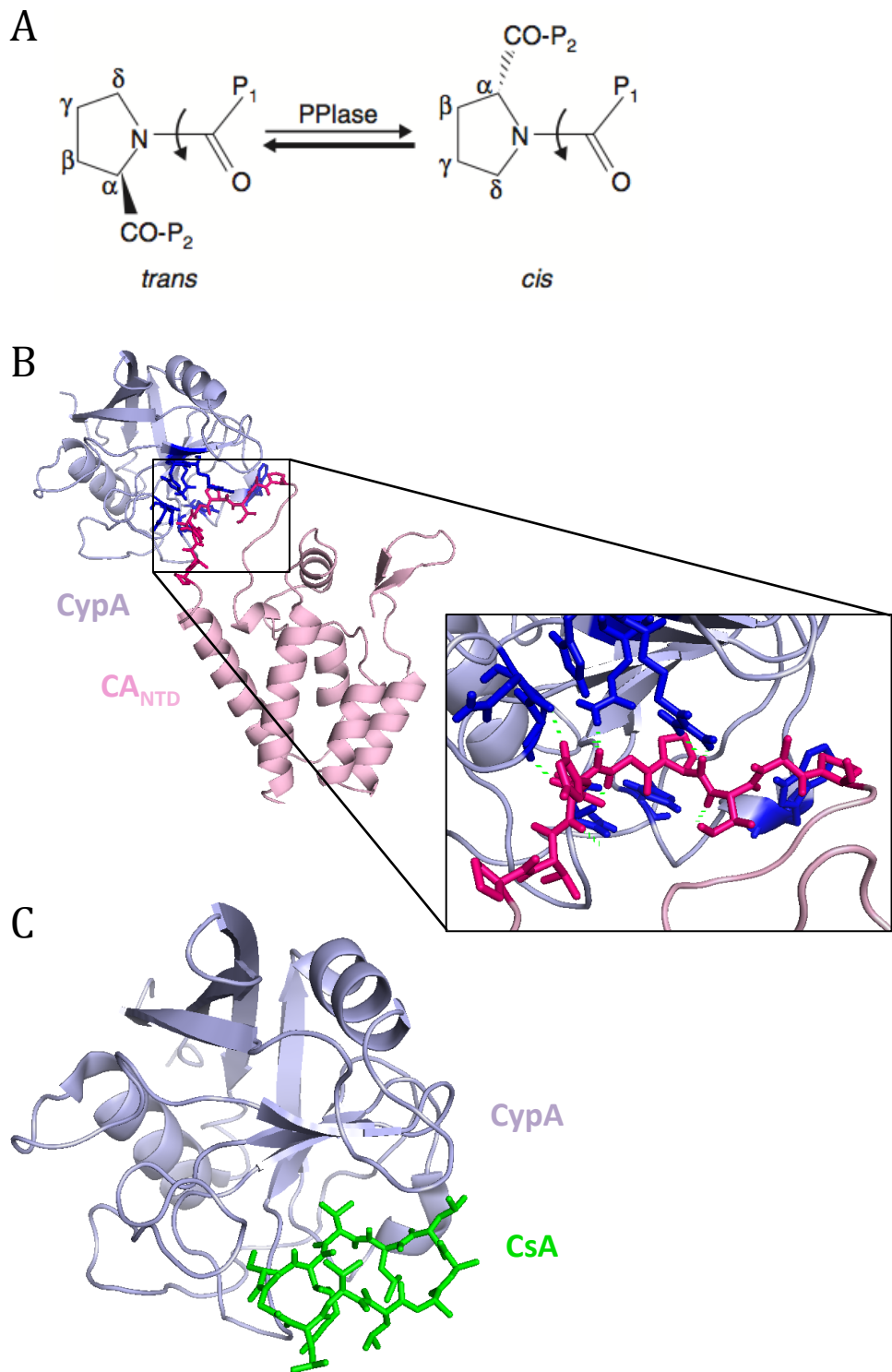
Figure 15. The interaction of CypA and HIV-1 CA

(A) *cis*/*trans* isomerisation of proline peptide bonds, as catalysed by peptidyl-prolyl isomerases like CypA. Image from Wang and Heitman 2005. **(B)** Crystal structure of CypA in purple bound to HIV-1 CA_{NTD} in pink (PDB 1AK4 from Gamble et al. 1996). Interacting residues (including the CypA-binding loop of CA) are highlighted in bold sticks and hydrogen bonds shown as green dashes in the magnified image. **(C)** Crystal structure of CypA in purple bound to CsA in green (PDB 1CWA from Mikol et al. 1993).

osteosarcoma viral oncogene homolog (Fos) and Jun proto-oncogene (Jun). This complex upregulates interleukin 2 (IL-2) transcription (280, 281). IL-2 then stimulates growth and differentiation of T cells and the production of other cytokines. CsA interferes with this pathway by forming a ternary complex with CypA and calcineurin (282-284), which prevents calcineurin from interacting with NFAT (285, 286).

1.5.2.2 *Target cell CypA enhances viral infectivity by interacting directly with CA*

HIV-1 Gag (p55) was originally found to interact with CypA and the closely related CypB through a yeast-2-hybrid screen of a human cDNA library, and the region of Gag responsible for mediating CypA interaction was found to be CA (p24) (287-289). Residues 85-93 of CA_{NTD} form a dynamic proline-rich loop (now commonly referred to as the CypA-binding loop), which binds to the catalytic site of CypA's PPIase domain (89) as shown in figure 15B. CA residues G89 and P90 are essential for interaction with CypA. The CA mutation G89V abolishes interaction with CypA and P90A significantly reduces interaction (290). P85, V86, H87, A88, and P93 also make energetically favourable contacts, and mutations of these residues reduce the affinity of CA for CypA (288, 290). CsA inhibits CA-CypA interaction by competing with CA for the PPIase pocket of CypA as shown in figure 15C (291). This region of CA is highly conserved in HIV-1 group M isolates and SIVcpz, but varies significantly in group O isolates, HIV-2, SIVagm, SIVmac, and SIVsm. As would be expected, divergence from this conserved sequence correlates with a reduction in CypA binding (292). The interaction between CA and CypA means it is specifically packaged into HIV-1 virions and so it was suggested that this packaged CypA was essential for viral infectivity (110, 288, 289, 293). However, Towers et al. (294) demonstrated that target cell CypA, as opposed to producer cell CypA, was required for HIV-1 infectivity in human cells (cell lines and primary cells).

1.5.2.3 The relationship between HIV-1 and CypA is cell type-dependent

Interestingly, the relationship between CA and CypA is cell-type dependent. WT HIV-1 is inhibited by CsA in cell lines such as Jurkat and TE671 and in primary peripheral blood lymphocytes (PBLs) but was not inhibited by CsA (in some cases CsA even slightly enhances infectivity) in other cell lines, such as H9 and HeLa (294-296). This cell-type dependent nature of CypA is emphasised by two CA mutants, A92E and G94D, which were created by Braaten *et al.* by serially passaging HIV-1 in Jurkat cells in the presence of CsA. Despite lying within the CypA-binding loop of CA, these mutations do not alter the affinity of CA for CypA (297). Intriguingly, their phenotype is cell-type dependent; they are CsA-resistant in 'permissive' cells such as Jurkat and TE671 but CsA-dependent in 'non-permissive' cells such as H9 and HeLa (296). Like CsA, CypA depletion or P90A or G89V mutation can rescue A92E and G94D in non-permissive cells by reducing the interaction of CypA with the virus (296). A study of six cell lines found that permissive cells had lower CypA levels than non-permissive cells (295). Ylinen *et al.* (298) subsequently showed that overexpression of CypA in permissive TE671 cells caused them to become non-permissive. They also showed that A92E or G94D infectivity could be rescued by addition of CsA up to 10hrs post-infection, suggesting that these mutants are inhibited because they are hyperstable (rather than because they are unstable and prematurely uncoat, a process that would presumably not be reversible after 10hrs). In light of these observations, it was hypothesised that CypA increases core stability and that A92E and G94D mutations confer increased core stability because they evolved to compensate for a lack of CypA (having been selected for in the presence of CsA).

In disagreement with these observations, a recent study of 27 different cell lines and 32 HeLa cell clones found no significant correlation between CypA expression and infectivity (112), calling into question this conclusion. Heterokaryons between non-permissive HeLa cells and permissive 293T cells revealed that the non-permissive phenotype is dominant (299). It was argued that if this phenotype were solely due to CypA expression, heterokaryons would have an intermediate phenotype. It has therefore also been proposed that a

CypA-dependent restriction factor may be present in (some, if not all) cell lines that are non-permissive to A92E and G94D. However, a recent genome-wide expression analysis failed to identify such a restriction factor (300).

Curiously, a titration of CsA revealed a biphasic inhibition curve of WT HIV-1 in most cell lines tested; low concentrations of CsA (<0.1μM) enhanced infectivity but higher concentrations inhibited the virus (301). This could be interpreted to mean that there is a weak CypA-dependent restriction factor present in most cells but that CypA is also a cofactor of HIV-1 infection. If so, a small decrease in CypA concentration increases infectivity by relieving HIV-1 from restriction, but a larger decrease in CypA concentration decreases infectivity because it is needed as a cofactor.

TRIM5α is an example of a known restriction factor that inhibits HIV-1 in a CypA-dependent manner (for details of TRIM5α and its mechanism of action see section 1.3.1). Isoforms of TRIM5α from Old World monkeys (OWMs) such as African green monkeys (agmTRIM5α) and rhesus macaques (rhTRIM5α) potently restrict HIV-1 in a CypA-dependent manner (294, 302, 303). In contrast, CypA does not enhance huTRIM5α-mediated restriction (294, 304) and has even been proposed to protect HIV-1 from restriction (294). Importantly, Shah and Aiken (300) found that huTRIM5α expression levels could not explain the differences in A92E permissivity between cells, suggesting that huTRIM5α is not the cause of this phenotype.

1.5.2.4 CypA may alter HIV-1 core stability

As discussed in section 1.5.2.2, CA residue P90 interacts directly with the PPIase active site of CypA (89). Nuclear magnetic resonance (NMR) spectroscopy has revealed that CypA catalyses the *cis-trans* isomerisation of the peptide bond within this proline residue (305), but the significance of this isomerisation has been disputed. Lammers, et al. (306) solved crystal structures of CA in complex with CypA or acetylated CypA (which is catalytically inactive) in which the P90 peptide bond adopted the *trans* or *cis* conformation, respectively. Despite this,

there were no substantial changes in CA structure outside the CypA-binding loop and the affinity of CA for CypA was not affected. It is therefore considered unlikely that the PPIase activity of CypA influences HIV-1 core stability.

Whilst CypA may not alter core stability through PPIase activity, it may do so through steric effects. Gamble, et al. (89) estimated that the CA-CypA ratio in a virion is 10:1 and suggested that this would sufficient to induce 'a series of minor dislocations' that could lead to a weakening of the lattice and thus disassembly. Biochemical studies of the effect of CypA upon core stability have produced contradicting results, which some reporting that CypA does indeed destabilise cores (90) and others that it stabilises cores (87, 112).

1.5.2.5 CypA determines the HIV-1 nuclear entry pathway and integration site selection

Schaller *et al.* (79) recently reported that that the sensitivity of HIV-1 to depletion of the nuclear entry cofactors discussed above is dictated by the interaction of CA with CypA. They found that inhibiting CA-CypA interaction – either by G89V or P90A mutation of CA, CypA depletion, or CsA treatment – caused HIV-1 to become insensitive to Nup358 depletion and significantly less sensitive to TNPO3 or Nup153 depletion. The insensitivity of G89V and P90A to Nup358 depletion could be explained by the fact that these mutations prevent CA-Nup358$_{Cyp}$ interaction. However, neither CypA depletion nor CsA treatment affects the affinity of CA for Nup358$_{Cyp}$ and so the insensitivity of WT HIV-1 to Nup358 depletion under these conditions must be due to the lack of CA-CypA interaction. These observations suggest that cytoplasmic interaction of the viral core with CypA can determine the downstream nuclear entry pathway taken by the virus.

Based upon these observations, one may predict that G89V or P90A mutation or CsA would reduce the average gene density surrounding HIV-1 integration sites, as has been shown for TNPO3 or Nup358 depletion. Surprisingly, the opposite is true. G89V mutation or CsA treatment increased the average gene density of

HIV-1 integration sites to increase from ~15genes/Mb to ~20genes/Mb (79). As discussed in section 1.2.9.6, it has been hypothesised that a very high gene density may be disadvantageous to the virus, although this remains to be investigated.

As well as being able to dictate dependence of HIV-1 upon downstream cofactors, CypA has been shown to dictate sensitivity of HIV-1 to MxB, a recently identified IFN-inducible HIV-1 restriction factor that inhibits HIV-1 nuclear entry (307, 308). MxB has been suggested to interact directly with HIV-1 cores and to prevent them from uncoating (117). Serial passage of HIV-1 in SupT1 cells overexpressing MxB lead to the outgrowth of MxB-resistant viruses, which were repeatedly found to carry the mutation A88T in CA. A88 lies within the CypA binding loop of CA and A88T is a known CypA binding mutant (309). Other CypA-binding mutations in CA such as P90A and G89V have also been found to confer MxB resistance, as has CsA (307-309). It therefore appears that CA-CypA interaction dictates the sensitivity of HIV-1 to downstream restriction factors as well as the dependence of HIV-1 upon downstream cofactors.

1.5.3 A role for CPSF6 in HIV-1 infection

1.5.3.1 *CPSF6 plays a role in 3' pre-mRNA processing and mature mRNA export*

Eukaryotic pre-mRNAs are post-transcriptionally processed to become mature mRNAs in two steps; endonucleases cleave the 3' end of the primary transcript at a CA dinucleotide in the UTR, and subsequently poly(A) polymerases (PAPs) polyadenylate the new 3' end. These two steps involve several *cis*-acting RNA elements and an estimated 80 *trans*-acting protein factors in humans (310). The components required for each step have been dissected using *in vitro* 3' processing reactions. Originally, five protein fractions with unique functions were identified, and each of these is now known to be a complex of multiple proteins. One of these complexes is cleavage factor Im (CFIm) (310, 311). UV cross-linking experiments revealed that CFIm can interact with RNA, and that it stabilises the interaction between RNA and the cleavage and polyadenylation

specificity factor (CPSF) (312), a complex which is involved in recognition of the poly(A) signal (PAS) (310). CFIm is a heterotetramer of two molecules of CPSF5 and two molecules of either CPSF6[68], CPSF6[72], or CPSF7 (312-315); the *CPSF6* gene encodes the two splice variants CPSF6[68] and CPSF6[72] (simply CPSF6 will be used when referring to both proteins), which are 68kDa and 72kDa respectively, whilst the *CPSF7* gene is a paralog of *CPSF6* (316). These three large subunits are likely redundant and CPSF6[68], which is by far the most abundant protein, is sufficient for *in vitro* 3' processing (313, 316).

Figure 16A shows the domain organisation of CPSF6 and CPSF5. CPSF6 comprises two C-terminal RNA recognition motifs (RRMs), a central proline-rich (pro-rich) domain, and an N-terminal arginine/serine-rich (RS) domain (313, 317). CPSF5 shows homology to the nucleoside diphosphate linked to x (Nudix) family of phosphohydrolases but lacks key catalytic residues. Instead, this region is thought to mediate protein-protein interactions with PAP (316). Crystallography has revealed that a dimer of CPSF5 forms a heterotetramer with two monomers of the RRM of CPSF6, as shown in figure 16B, and that this interaction required at least a 9-mer of RNA bound to CPSF5 (315). As shown in figure 16C, each CPSF5 molecule binds directly to a UGUA motif and the CPSF6 molecules are thought to enhance CPSF5-RNA interactions and to facilitate looping of the RNA (315, 317, 318).

In addition to its role in processing of pre-mRNA at the PAS, CPSF6 has been found to fractionate with spliceosomes and to interact with splicing factors, suggesting that it may coordinate splicing and 3' processing (316, 319). A yeast-2-hybrid screen using the arginine/serine-rich domain (RS domain) of CPSF6 as bait identified strong interactions with several splicing factors belonging to the serine/arginine-rich protein family (SR proteins) (317). Further supporting a role for CPSF5 and/or CPSF6 in splicing regulation is the finding that dephosphorylation of the CFIm complex inhibits pre-mRNA cleavage (316, 320). Immunofluorescence studies have shown that endogenous CPSF6 is exclusively nuclear at steady state. Whilst it is found in a diffuse pattern
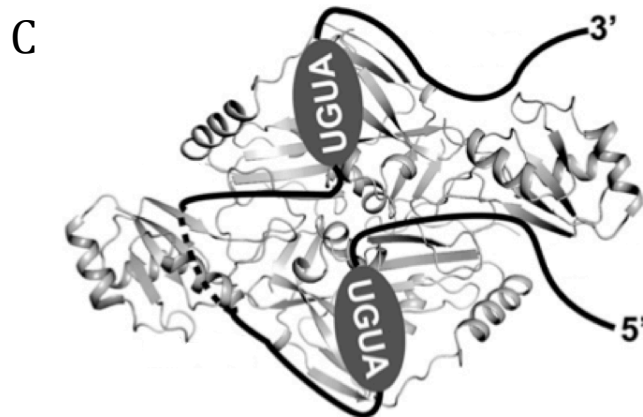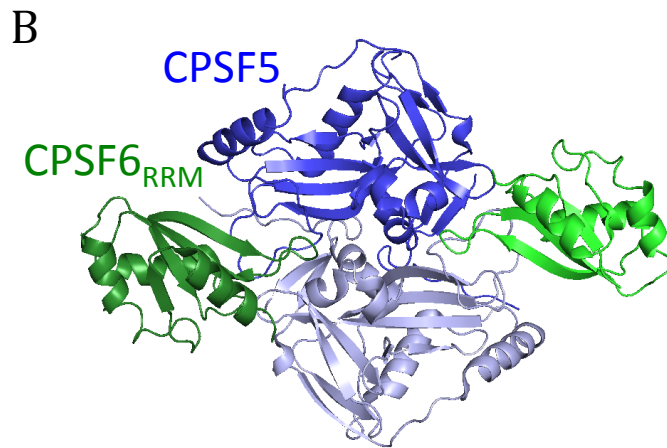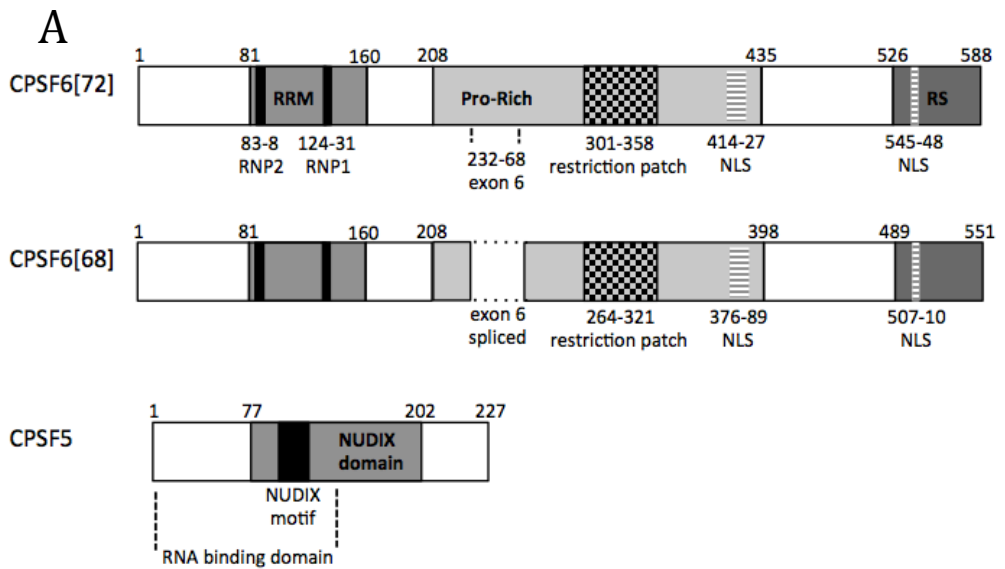
Figure 16. The structures of CPSF6, CPSF5, and CFIm
(A) Schematic diagram of the domains of CPSF6[72] and [68] and CPSF5. **(B)** A crystal structure of part of CFIm. Shows a heterotetramer comprising a dimer of CPSF5 in blue bound by two CPSF6 RRMs in green (PDB 32QS from Yang et al. 2011). **(C)** Model of how CFIm in grey interacts with RNA in black. Each molecule of CPSF5 interacts directly with a UGUA motif in RNA. CPSF6 enhances CPSF5-RNA interaction and aids bending/looping of the RNA. Image adapted from Yang et al. 2011.

throughout the nucleus, it is also enriched in nuclear speckles and paraspeckles. Thus, it colocalises with RNA, RNAPII, and splicing component 35 (SC35), suggesting that splicing, 3' processing, and transcription may all occur consecutively or even simultaneously in one location (321). Immuno-EM has been used to show that CPSF6 localises to perichromatin fibrils, which are sites of pre-mRNA processing found adjacent to sites of active transcription, only when active transcription is occurring (321).

Despite exclusively nuclear staining of steady-state CPSF6 (321), several proteins shown to interact with CPSF6 are thought to shuttle between the nucleus and the cytoplasm (317). Therefore, Ruepp, et al. (322) used heterokaryon experiments to test their hypothesis that it was CPSF6 regulating their movements. They expressed hemagglutinin-tagged (HA-tagged) CPSF6 in HeLa cells and subsequently fused them to 3T3 cells using PEG (polyethylene glycol). By 2 hours post-fusion HA-CPSF6 was present in both the HeLa and 3T3 nuclei of all heterokaryons examined whereas the nucleus-restricted negative control protein (heterogeneous nuclear RNP-C) had not move between nuclei, unequivocally demonstrating the ability of CPSF6 to shuttle between the nucleus and cytoplasm (322). Shuttling of CPSF5 tagged with yellow fluorescent protein (YFP) was found to be much less efficient, but was significantly enhanced by co-expression of GFP-tagged CPSF6. Export of CPSF6 from the HeLa nuclei was prevented by the addition of Actinomycin D, which inhibits transcription by preventing elongation of mRNA by RNA polymerase, implying that CPSF6 export may be dependent upon mRNA traffic. Supporting this is the observation that the N-terminal region (the RRM) of CPSF6 interacts directly with nuclear RNA export factor 1 (NXF1). Importantly, CPSF6 was also shown to stimulate mRNA export through an elegant assay; cells were transfected with a plasmid expressing a luciferase mRNA that contains an inefficiently spliced intron and 6 binding sites for the bacteriophage MS2 coat protein. Normally, this mRNA is not spliced and so is not exported or expressed. However, in the presence of RNA export factor (REF) (which interacts with NXF1) fused to the bacteriophage MS2 coat protein, the mRNA can be exported despite being unspliced and is thus expressed. CPSF6 was able to substitute for REF in this

assay, revealing its ability to promote mRNA export, likely through interaction with NXF1 (322). Subsequent fractionation experiments suggested that CPSF6 is associated with cytoplasmic messenger ribonucleoprotein complexes, but not with translating ribosomes, suggesting it may be involved in initiation of translation but dissociates before elongation begins (322). From these studies we infer that CPSF6 shuttles in and out of the nucleus, spending little time in the cytoplasm, which may explain why immunofluorescence studies do not detect significant cytoplasmic CPSF6 staining.

1.5.3.2 C-terminal truncations of CPSF6 inhibit HIV-1

CPSF6 was first implicated in HIV-1 infection when Lee *et al.* (259) reported the results of a screen of a 3T3 cell cDNA library for murine restriction factors of HIV-1. They identified mCPSF6[72]-358, a form of murine CPSF6[72] that was C-terminally truncated at residue 358, which was able to restrict HIV-1 infection after reverse transcription but prior to nuclear entry (prior to 2-LTR circle formation). It was shown to restrict both VSV-G pseudotyped HIV-1 vector and replication-competent HIV-1 in both murine (3T3) and human (HeLa) cell lines, as well as in primary human T cells. Excess virus or virus-like particles (VLPs), which contain no genome, were able to saturate restriction by mCPSF6[72]-358, suggesting a virus-specific mechanism of action as opposed to disruption of a cellular process required for viral infection. In further support of this, restriction was found to be specific to primate Lentiviruses (HIV-1, HIV-2, SIVmac, and SIVmne) and did not impact the infectivity of the γ-retrovirus Murine Leukaemia Virus (MLV).

In order to determine the viral requirements for restriction, Lee *et al.* serially passaged the replication-competent HIV-1$_{NL43/BaL}$ in HUT-R5 cells expressing mCPSF6[72]-358 and sequenced the outgrowing virus for mutations. The mutation N74D in CA repeatedly arose and allowed complete escape from mCPSF6[72]-358, suggesting that this restriction factor inhibits the virus through a specific interaction with CA. Residue N74 is highly conserved in primate Lentiviruses and lies within helix 4 of CA$_{NTD}$, which is predicted to be

accessible in the context of a hexamer. It was confirmed that mCPSF6[72]-358 from cell lysates immunoprecipitates with CA-NC tubes *in vitro* and that N74D mutation significantly reduces the level of precipitation (259). This could represent a direct or indirect interaction between mCPSF6[72]-358 and CA. As well as identifying CA residues essential for its interaction with mCPSF6[72]-358, Lee *et al.* narrowed down the region of CPSF6 required for restriction. They found that mCPSF6[72]-300 could not inhibit HIV-1 and interacted with CA-NC significantly more weakly than mCPSF6[72]-358, suggesting that residues 300-358 are necessary for interaction with CA and for inhibition of the virus.

1.5.3.3 A role for full-length CPSF6 in HIV-1 infection

Importantly, mCPSF6[72]-358 is not a naturally occurring protein in mice or humans. Instead, this truncated form of CPSF6 is an artefact of the cDNA library screening methodology used, which used random primers to amplify DNA. Lee *et al.* therefore tested whether full-length human CPSF6[72] was able to inhibit HIV-1 and found that it caused a 2-fold reduction in titre. They also showed that RNAi-mediated depletion of endogenous CPSF6 in HeLa cells lead to a ~20% increase in infectivity. They considered these effects to be significant and suggested that endogenous CPSF6 may be an HIV-1 restriction factor.

They acknowledged that C-terminally truncated forms of CPSF6 were much more potent inhibitors and sought to determine why. They noted that the truncated C-terminus of CPSF6 has been reported to contain a classical NLS (317). We confirmed that the RS domain of CPSF6 contains a classical monopartite NLS with the consensus of three or four R/K residues and one H/P residue ($_{507}$RRHK$_{510}$ in CPSF6[68]) using PSORT II (http://psort.hgc.jp/form2.html). Lee *et al.* showed that C-terminal truncating CPSF6[68] or [72] just before this NLS (creating hCPSF6[68]-526 or mCPSF6[72]-526) rendered it antiviral. Through fractionation of 3T3 cells they showed that, whilst mCPSF6[68] was primarily nuclear, these NLS-deleted forms of CPSF6 were present in both the cytoplasmic and nuclear fractions.

CPSF6 has also been reported to interact with karyopherin β2 (323), which binds cargo through a non-classical bipartite NLSs with the consensus sequences ϕG/A/S$\phi\phi$X$_{(11-13)}$PY and R/K/HX$_{(2-5)}$PYC (ϕ denotes a hydrophobic residue). CPSF6 has one such NLS in its pro-rich domain (CPSF6[68] $_{376}$RGPP$_{379}$...$_{385}$RPPPY$_{389}$) (324). mCPSF6[72]-526 contains this predicted NLS, but mCPSF6[72]-358 does not. The fact that both of these proteins were primarily cytoplasmic suggests that removal of the classical NLS is sufficient to mislocalise the protein.

Based upon their observations Lee *et al.* hypothesised that endogenous CPSF6 weakly restricts HIV-1 nuclear entry but may be a more potent inhibitor in circumstances where its concentration in the cytoplasm is increased, although they did not speculate as to what circumstances may cause a change in CPSF6 localisation. They postulated that CPSF6 restricts HIV-1 indirectly by preventing interaction of CA with a cofactor of nuclear entry.

1.5.3.4   *N74D(CA) enters the nucleus through a different pathway from WT HIV-1 and cannot replicate in MDMs*

In order to test their hypothesis that mCPSF6[72]-358 prevents interaction of HIV-1 CA with a cofactor, Lee *et al.* examined the sensitivity of the mCPSF6[72]-358 escape mutant N74D to depletion of known HIV-1 cofactors. Intriguingly, they found that N74D mutation makes HIV-1 insensitive to depletion of the cofactors TNPO3, Nup358, and Nup153 but causes it to become sensitive to depletion of the nucleoporins Nup85 and Nup155. They suggested that N74D is able to avoid interaction with the restriction factor CPSF6 by using a different nuclear entry pathway. Schaller *et al.* (79) recapitulated the finding that N74D is insensitive to TNPO3 or Nup358 depletion and less sensitive to Nup153 depletion.

In accordance with the finding that TNPO3 or Nup358 depletion reduces the gene-density surrounding HIV-1 integration sites from ~21 to ~11genes/Mb

(159) (discussed in sections 1.5.1.1 and 1.5.1.2), they also found that N74D mutation caused HIV-1 to integrate into regions of a gene density equivalent to a random site in the genome (~7genes/Mb). In other words, N74D mutation caused HIV-1 to lose its ability to selectively integrate into gene-enriched regions. It is intriguing that N74D and the CypA-binding mutant G89V have such opposing effects upon HIV-1 integration site selection given that they both exhibit reduced dependence upon nuclear entry cofactors. Together these data suggest that, in HeLa cells at least, HIV-1 is capable to utilising multiple different nuclear entry pathways that result in different integration site targeting. Although N74D is reportedly as infectious as WT HIV-1 in HeLa cells, Schaller *et al.* found that it was unable to replicate in primary monocyte-derived macrophages (MDMs) (79). This suggested that the nuclear entry pathways taken by HIV-1 in cell lines such as HeLa differs from that taken in primary cells such as MDMs.

1.6 Aims of this study

As detailed in chapter 3, our initial aim was to determine whether CPSF6 does indeed play a role in HIV-1 infection, as had been suggested by Lee *et al.* (259). In our hands, CPSF6 did not affect HIV-1 infectivity of HeLa cells. However, we provide evidence that CPSF6 (but not its binding partner CPSF5) influences the dependence of the virus upon the interaction of the virus with downstream nuclear entry cofactors and upon integration site targeting. We interpret these observations to mean that HIV-1 can take several different nuclear entry pathways of equal efficiency and that the CA-CPSF6 interaction can determine which pathway is taken.

As discussed in chapter 4, we examined clinically relevant CA mutations associated with escape from B*57/27$^+$ CTL responses that reportedly alter the relationship of HIV-1 with CypA (325, 326). We recapitulated the reported observations and hypothesised that these mutants may also interact differently with other early cofactors of HIV-1 infection (specifically those discussed in chapter 3). Indeed, we found that the B*27-associated CTL escape mutation R132K causes HIV-1 to enter the nucleus via a different pathway than WT HIV-1. We found that the compensatory mutation S41A largely restores the WT phenotype but identified a few key phenotypic differences that suggest other unknown cofactors are likely involved in HIV-1 nuclear entry and integration site selection. These observations suggest that adhering to the CPSF6-dependent nuclear entry pathway described in chapter 3 is important *in vivo*.

In chapter 5 we discuss the recent discovery of antiviral compounds that interact with CA in the same pocket as CPSF6 (111, 327). We examine the mechanism of action of these compounds, with the aim of better understanding how different molecules interacting with the same pocket can have different effects upon the virus. We provide evidence that these antiviral compounds inhibit the virus by a reversible mechanism and that the concentrations required to displace CPSF6 are lower than those required to inhibit the virus, an observation that separates the two phenotypes of these compounds. In agreement with this, we were able to rationally design a compound that inhibits

CA-CPSF6 interaction without inhibiting the virus, thereby altering the nuclear entry pathway of the virus.

Throughout this study, our aim has been to characterise the role of CPSF6 in HIV-1 infection, both *in vitro* and *in vivo*, and also to determine whether CPSF6 is a viable antiviral drug target. We provide evidence that CPSF6 is a cofactor of HIV-1 infection, albeit not in a conventional manner, and that drugging the CA-CPSF6 interaction may be able to inhibit the virus through more than one mechanism.

2 Chapter 2. Methods and materials

2.1 Plasmids and preparation

2.1.1 Plasmids

The details of the plasmids used in this study can be found in Table 1.

Table 1: Plasmids used in this study

Plasmid name	Plasmid category	Expresses	Details	Ref/ supplier
pMD.G	Envelope	VSV-G protein	CMV promoter	(328)
p8.91Ex	Packaging	HIV-1 gag-pol, tat, rev		(329)
pCIG3-B	Packaging	B-MLV gag-pol		(330)
pCMVi	Packaging	MoMLV gag-pol		(331)
pCSGW	Transfer (SIN)	eGFP	HIV-1 LTRs	(332)
pCSLW	Transfer (SIN)	Luciferase (firefly)	HIV-1 LTRs	(333)
pCNCG	Transfer	eGFP	MLV LTRs	(333)
RNAi Ready pSIREN RetroQ	Transfer	shRNA Puromycin^r	MoMLV LTRs CMV enhancer Human U6 promoter	Clontech
pEXN	Transfer	HA-tagged transgenes G418^r	MoMLV LTRs Derivative of pLNCX2	(334)
pBru3oriGFP3 ΔEnv	HIV-1 ΔEnv	HIV-1 LAI ΔEnv eGFP in place of Nef		(335)

Packaging plasmids express proteins required for virus assembly and packaging of the viral genome. Transfer plasmids express the viral genome that will be packaged into virions. This genome encodes the transgene/s that will be expressed by the virus. Envelope plasmids express the envelope protein that will assemble onto the surface of virions. VSV = Vesicular Stomatitis Virus. CMV = Cytomegalovirus. MLV= Murine Leukaemia Virus. MoMLV = Moloney MLV. SIN = self-inactivating. SIN transfer vectors have a deletion in the U3 region of their 3' LTR which, during reverse transcription, is copied to produce the 5' LTR. This deletion means they lack viral promoter/enhancer activity in the LTR and new genomes cannot be synthesised. They therefore have another internal promoter, the Spleen Focus-Forming Virus (SFFV) LTR promoter, to allow expression of their transgene. eGFP = enhanced GFP

2.1.2 Preparing HB101 chemically competent bacteria

HB101 (F$^-$ mcrB mrr hsdS20(r_B^- m_B^-) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(Sm^R) glnV44 λ$^-$) as described in (336) were used to inoculate 5ml of lysogeny broth (LB) (tryptone 10g/L, yeast extract 5g/L, NaCl 5g/L, sterilized by autoclaving). This culture was incubated, shaking, at 37°C for approximately 18 hours. 1ml of this culture was then used to inoculate 200ml LB. This culture was incubated, shaking, at 30°C until an OD595 of 0.5-0.6 was

achieved (as measured using a WPA UV1101 spectrophotometer (Biotech Jencons)). The culture was divided into 4x50ml falcon tubes and chilled on ice for 10mins. Cultures were centrifuged at 4°C and 3000rpm, for 10mins. The supernatant was discarded, and the bacterial pellets pooled and gently resuspended in a total of 20ml TFB1 buffer (30mM KAc, 100mM RbCl, 10mM CaCl, 50mM MnCl, 15% glycerol (in dH_2O), 0.45μm filter sterilized), before chilling on ice for 5mins. This was again centrifuged at 4°C and 3000rpm, for 10mins. The supernatant was discarded, and the bacterial pellet gently resuspended in 2ml TFB2 buffer (10mM PIPES, 75mM $CaCl_2$, 10mM RbCl, 15% glycerol (in dH_2O), 0.45μm filter sterilized), before chilling on ice for 10mins. The bacteria were aliquoted (50μl each) before being snap frozen on dry ice. The competent bacteria were stored at -80°C. Competency was tested by transforming 50μl HB101 with 10pg of pUC19; over 100 colonies indicated a high degree of competency.

2.1.3 HB101 transformation and selection

50μl aliquots of chemically competent HB101 cells were thawed for 10mins on ice, before being added to plasmid DNA (≤100ng and ≤5μl), and then chilled for a further 30mins on ice. Bacteria were then heated shocked at 42°C for 45secs, and chilled for a further 2mins on ice. Bacteria transformed with a plasmid encoding Ampicillin resistance were plated onto selective agar immediately after transformation. Bacteria transformed with a plasmid encoding Kanamycin resistance were 'recovered' by growing in LB at 37°C for 1hr before plating into selective agar. Solid agar was made by dissolving 17.5g agar (VWR) in 500ml dH_2O and autoclaving on a liquid cycle before storing them at 4°C. Selective agar plates were made by melting this solid agar, adding the appropriate antibiotic for selection (Ampicillin (Calbiochem) used at 100μg/ml, or Kanamycin (Sigma Aldrich) used at 30μg/ml), pouring it into petri dishes and cooling them at room temperature, before storing them at 4°C. After transformed bacteria were plated, they were incubated upside down at 37°C for approximately 18hrs.

2.1.4   Plasmid miniprep

Single colonies were picked from agar plates, and used to inoculate 5ml LB containing the appropriate selection. Cultures were incubated, shaking, at 37°C for approximately 18 hours. Bacteria were pelleted by centrifugation at 3000rpm for 3mins, and the supernatant discarded. Plasmid DNA was extracted using QIAprep Spin Miniprep Kit (Qiagen). Plasmid DNA concentration was measured using Nanodrop ND-1000 (Thermo Fischer Scientific Inc.). Concentrations between 0.2-1µg/µl were obtained.

2.1.5   Plasmid midiprep

Single colonies were picked from agar plates, and used to inoculate 5ml LB containing the appropriate selection. Cultures were incubated, shaking, at 37°C for approximately 8 hours. 200µl of this culture was then used to inoculate 200ml LB, which was incubated, shaking, at 37°C for approximately 18 hours. Cultures were centrifuged at 4000g, at 4°C, for 20mins, and the supernatant discarded. Plasmid DNA was extracted using QIAfilter Plasmid Midi Kit (Qiagen). Concentrations between 0.8-2µg/µl were obtained.

2.2   Site directed mutagenesis (SDM)

SDM was used throughout this study to introduce point mutations into plasmids.

2.2.1   SDM primer design

SDM primers were designed to – where possible - be 30-45bp long, to overlap with each other by at least 20-30bp, to have 40-70% GC content, to begin and end in a G/C, to have a melting temperature (T_m) between 55 and 80°C and to have similar T_ms to each other, a maximum of 3 mismatches to be introduced, and for the mutation/s to be positioned toward the 3' end of the primer. For primer sequences used in this study, see Table 2.

Table 2: SDM primer sequences

CA mutation	Plasmid mutated	Primer ID	Sequence (5'-3')
T110N		JCR45	GAGATCTGGCTGAACATCTTAAGACAGCAG
		JCR46	GTTCAGCCAGATCTCTTACCTGTCCTATAATTTTC
H87Q		JCR47	GTGCATCCAGTGCAGGCAGGGCCTATTGCAC
		JCR48	GCCCTGCCTGCACTGGATGCACTCTATCCC
G116A		JCR55	CAAATAGCCTGGATGACACATAATCCACCTATC
	p8.91Ex	JCR56	CATCCAGGCTATTTGTTCCTGAAGGGTACTAG
M96I		JCR51	GGCCAGATCAGAGAACCAAGGGGAAGTGAC
		JCR52	GTTCTCTGATCTGGCCTGGTGCAATAGGCCC
I91V		JCR49	CAGGGCCTGTGGCACCAGGCCAGATGAGAGAAC
		JCR50	CTGGTGCCACAGGCCCTGCATGCACTGG
L136M		JCR77	GGAGAAATCTATAAAAGATGGATAATCATGGGATTAAATAAAATAG
		JCR75	GGGCTATACATTCTTACTATTTTATTTAATCCCATGATTATCC
R132K	p8.91Ex L136M	JCR76	GTAGGAGAAATCTATAAAAAATGGATAATC
		AF356	CCATTTTTTATAGATTTCTCCTACTGGG
S41A	p8.91Ex R132K, L136M	AF313	CATGTTTGCCGCATTATCAGAAGGAGCCACCCC
		AF314	GATAATGCGGCAAACATGGGTATCACTTCTGGGC
A/V/T41S	pBru3ori GFP3ΔEnv viruses	JCR94	ATACCCATGTTTTCAGCATTATCAGAAGGAGCCACCCC
		JCR95	CTTGTGGGGTGGCTCCTTCTGATAATGCTGAAAACATGGG

2.2.2 SDM PCR setup

SDM PCRs were catalysed using Pfu Turbo DNA Polymerase (Agilent). For details of reaction setup, see Table 3. For cycling parameters, see Table 4. The annealing temperature of the SDM PCR was adjusted between 50-60°C when trouble shooting. For each PCR, a negative control that contained no Pfu Turbo was included. From each 50μl PCR, 6μl was run on an agarose gel containing ethidium bromide (see section 2.3.2); if the PCR was successful, DNA can be visualised under U.V. light.

Table 3: SDM PCR reagents

Component	Plasmid DNA <10kb		Plasmid DNA >10kb	
Pfu Turbo DNA Polymerase 2.5U/μl	2.5U	1μl	5U	2μl
Pfu Turbo 10x Buffer	1x	5μl	1x	5μl
dNTPs (Promega) 25mM each	250μM	5μl	500μM	10μl
Primers (Sigma) 100μM	2μM	1μl	4μM	2μl
Plasmid Template	50-100ng		100-200ng	
dH$_2$O		To total		To total
Final volume	50μl		50μl	

Table 4: SDM PCR cycling parameters

Step	Plasmid DNA <10kb			Plasmid DNA >10kb		
	Temp (°C)	Time (secs)	No. of Cycles	Temp (°C)	Time (secs)	No. of Cycles
Initial denaturing	95	30	1	92	30s	1
Denaturing	95	30	12/16/18 for 1/2/3 mutations	92	30	12/16/18 for 1/2/3 mutations
Annealing	55	60		55	60	
Extension	72	60/kb		68	120/kb	
Final extension	72	120/kb	1	68	240/kb	1

2.2.3 DpnI digestion and PCR cleanup

44µl of each completed PCR was added to 1µl DpnI (NEB) and 5µl Buffer 4 (NEB) and incubated at 37°C for 2hrs. DpnI is only able to digest methylated DNA, meaning that plasmid DNA produced by bacteria was digested whilst plasmid DNA produced by the PCR remained intact. The DpnI digested PCR was cleaned up using QIAquick PCR purification kit (Qiagen).

2.2.4 SDM product transformation and selection of successfuly mutated clones

After DpnI digestion and reaction cleanup, 5µl was transformed into HB101 competent bacteria and grown on selective agar (see section 2.1.3). If the negative control PCR produced no colonies, the DpnI digest was considered successful. Colonies were picked, cultured, and the plasmid prepared as in section 2.1.4. To identify successfully mutated clones, plasmids were sent for sequencing (Beckman).

2.3 Cloning and subcloning

After SDM of p8.91Ex (or its derivatives), a small region of the successfully mutated clone was subcloned back into the parental plasmid, and the subcloned region sequenced. This was to ensure that any mutations unintentionally induced by Pfu Turbo, which has an error rate of 1 per 1.3×10^{-6} (equivalent to 26% of 10kb products after 18 cycles), were removed. CA mutations made in a p8.91Ex background were subcloned between NotI and BglII. Similarly, after SDM of pBru3oriGFP3ΔEnv (or its derivatives), we subcloned between BssHII

and ApaI. DNA sequences designed to produce shRNAs (see section 2.7.1) were cloned into RNAi Ready pSIREN RetroQ (Clontech) between BamHI and EcoRI.

2.3.1   Restriction enzyme digestion

Promega and NEB supplied restriction enzymes. 3-5µg of plasmid DNA was digested in a total volume of 50µl for use in cloning, whereas approximately 500ng was digested in a total volume of 20µl for analytical purposes. Digests were carried out in the presence of bovine serum albumin (BSA, supplied as 100x) and the buffer (supplied as 10x) indicated by the manufacturer. Digests were carried out at the temperature indicated by the manufacturer, and for the appropriate amount of time given the concentration of the enzyme and the amount of DNA being digested.

2.3.2   Agarose gel preparation and running

1% agarose gels were made by dissolving 1g of powdered agarose (Sigma) in 100ml of 1x TAE (200ml 50x TAE (Tris acetate EDTA, VWR) made to 10L with dH$_2$O). Ethidium bromide was added to a final concentration of 0.2µg/ml (Sigma, supplied at 10mg/ml). DNA samples were run alongside GeneRuler 1kb DNA ladder (Thermo Scientific), using the DNA loading dye provided with the ladder. DNA bands were visualised by UV transillumination (UVP BioDoc-IT).

2.3.3   DNA extraction from agarose gels

DNA bands visualised on an agarose gel were cut out using a scalpel, and the DNA extracted using QIAquick Gel Extraction Kit or MinElute Gel Extraction Kit (both Qiagen) for products above and below 4kb, respectively. After elution, DNA was kept on ice at all times. DNA concentration was measured using the Nandrop ND-1000.

2.3.4 DNA ligation

Ligations were performed at vector:insert molar ratios of 1:3 and 1:10 (other ratios were tested when troubleshooting). Using 20-100ng of vector DNA, the required amounts of insert DNA was calculated using the following equation;

$$\frac{\text{ng of vector x kb of insert}}{\text{kb of vector}} \text{ x ratio = ng insert}$$

A vector-only negative control was also included. Ligations were carried out in a total volume of 20µl, using 1µl of T4 DNA Ligase (Fermentas) and the provided 10x buffer at a final 1x concentration. Ligations were incubated at 22°C for 1 hour (4°C for 16 hours was also tested when troubleshooting). Ligations were then transformed into HB101 competent bacteria and grown on solid agar containing the appropriate selection (see section 2.1.3). If the vector-only negative control produced no colonies this indicated that the digestion by both enzymes was complete and that no self-ligation of the vector had occurred. Colonies were picked, cultured, and the plasmid prepared as in section 2.1.4. To identify successfully mutated clones, plasmids were sent for sequencing (Beckman).

2.4 Mammalian cell culture

2.4.1 Mammalian cell line culture

The mammalian cell lines used in this study are detailed in Table 5. All cultures were supplemented with penicillin/streptomycin (Gibco Invitrogen, used at 100U/ml) unless otherwise stated. All cultures were incubated at 37°C and 5% CO_2 except for HEK 293T cells, which were incubated at 10% CO_2.

Table 5: Mammalian Cell Lines

Name	Cell Type	Culture Medium	Characteristics
HeLa	Cervical epithelial cells from human adult with cervical adenocarcinoma	DMEM 10% FBS	Adherent
HEK 293T	Kidney cells constitutively expressing Simian Virus 40 (SV40) large T antigen from human foetus	DMEM 15% FBS	Adherent. Highly transfectable.
TE671	Human embryo medulloblastoma cells	DMEM 10% FBS	Adherent
THP-1	Monocytes from human child with acute monocytic leukeamia	RPMI 10% FBS	Suspension
CEM	T lymphocytes from human child with acute lymphoblastic leukemia	RPMI 10% FBS	Suspension
Jurkat	T lymphocytes from human adult with acute T cell leukemia	RPMI 10% FBS	Suspension
CRFK	Feline kidney cortex cells	DMEM 10% FBS	Adherent

DMEM = Dulbecco's Modified Eagle Medium +L-Glutamine (Gibco Invitrogen). RPMI = Roswell Park Memorial Institute Medium 1640 (Gibco Invitrogen). FBS = Foetal Bovine Serum (BioSera).

2.4.2 Cell culture reagents

When washing cells, phosphate buffered saline (1x PBS –Ca –Mg (Lonza)) was used at 10ml/100mm dish. When removing adherent cells from culture plates, Trypsin 0.25% +EDTA (1x, Gibco Invitrogen) was used at 3ml/100mm dish. For storage, cells were pelleted by centrifugation at 3000rpm for 3mins and the supernatant discarded before resuspending in FBS+10% dimethyl sulphoxide (DMSO, Sigma). Cells were aliquoted into cryovials (at a density of $\geq 5 \times 10^6$ cells/ml and at least 500μl/cryovial) before being slowly cooled in a Mr Frosty (Thermo Scientific) container at -80°C. After 24hrs, cells could be transferred to liquid nitrogen for long term storage.

2.4.3 Primary CD4⁺ T cell culture

Primary CD4+ T cells were isolated and activated by Marie-Theresa Rodriguez-Plata as follows; Buffy coat residues were supplied by the National Blood Service and/or from healthy volunteers. Peripheral blood mononuclear cells (PBMCs) were obtained by Lymphoprep (Stemcell Technologies) density gradient centrifugation of heparin-treated blood. CD4⁺ T cells were purified from PBMCs with the indirect magnetic labeling system MACS (Miltenyi Biotec) according to manufacturer's instructions. Cells were cultured at a concentration of 2×10^6 cells/ml in RPMI supplemented with 10% heat-inactivated human

serum containing 10U/ml IL-2 (Roche Applied Science) and 3μg/ml phytohemagglutinin (PHA, SigmaAldrich). Cells were maintained at 37ºC in 5% CO2 in a humidified incubator.

2.5 Virus/vector production

2.5.1 Transfection and virus/vector collection

100mm plates of HEK 293T cells were transfected at 60-80% confluency. Per 100mm plate, for 3-plasmid transfections, the following was mixed and made to ≥15μl with TE (10mM TRIS pH8, 1mM EDTA); 1μg of packaging plasmid, 1μg of envelope plasmid, and 1.5μg of SIN transfer plasmid. For 2-plasmid transfections, 2.5μg of packaging/SIN transfer plasmid and 1μg of envelope plasmid was used. Per 100mm plate, for 1-plasmid transfections, 3.5μg of plasmid was made to ≥15μl with TE. Per 100mm plate, 10μl FuGene 6 (Roche) was added to 200μl OptiMEM Serum Free Media (Gibco Invitrogen), mixed by inverting, and added to the previously prepared DNA. Solutions were incubated at room temperature for 15mins. 293T cells were provided with 8ml fresh media prior to transfection. DNA/OptiMEM/FuGene mixtures were added to cells dropwise, whilst swirling the plate to ensure even distribution. At 48, 72, and 96hrs post-transfection, virus-containing supernatants were collected, 0.45μm filter sterilized, aliquoted, and stored at -80°C.

Table 6: Three-plasmid transfections

Vector type	Packaging plasmid	Transfer plasmid	Envelope plasmid
HIV-1(GFP)	p8.91Ex	pCSGW	pMDG
HIV-1(Luc)	P8.91Ex	pCSLW	pMDG
B-MLV(GFP)	pCIG3-B	pCNCG	pMDG
HIV-1 LAI ΔEnv GFP	pBru3oriGFP3ΔEnv		pMDG
Transgene-expressing	CMVi	pEXN	pMDG
shRNA-expressing	CMVi	RNAi Ready pSIREN RetroQ	pMDG

2.5.2 DNase treatment of virus

Plasmid contaminants of viral supernatants were removed by addition of DNase I (lyophilised bovine pancreas DNase I (Affymetrix), 6.110U/mg reconstituted to 1U/µl, used at 70U/ml virus) and DNase buffer (10x 400mM Tris HCl, 100mM $MgSO_4$, 10mM $CaCl_2$, pH8.0, 0.22µm filter sterilised, used at 1x) for 2hrs at 37°C.

2.5.3 Quantification of virus by RT enzyme-linked immunosorbent assay (ELISA)

Viral titres were standardised using the Colorimetric Reverse Transcriptase Assay (Roche). This ELISA involves a serial dilution of 0-2ng/well recombinant HIV-1 RT (for creation of a calibration curve), lysis of viral samples, an *in vitro* reverse transcription reaction that incorporates digoxigenin (DIG) and biotin-labeled nucleotides into reverse transcripts, attachment of reverse transcripts to microplate wells precoated in streptavidin, labeling of reverse transcripts with anti-DIG-peroxidase (POD) antibody, cleavage of a POD substrate, and measurement of the coloured reaction product using a Multiskan FC Microplate Photometer (ThermoScientific). Values between 20-200ng/µl were obtained. All samples were tested in duplicate.

2.6 Mammalian cell infections

All Lentiviruses used in this study are SIN vectors, meaning they can only complete a single-round of infection and express no proteins other than the transgene encoded by the transfer plasmid.

2.6.1 Infection for flow cytometry or luciferase assay

When using adherent cell lines, cells were seeded in 24 well plates at $5x10^4$/ml and 500µl/well. 24 hours later, media was discarded and replaced with 250µl media containing the appropriate concentration of virus and 8µg/ml Polybrene (PB, hexadimethrine bromide, Sigma). 48-72 hours later, cells were thoroughly

trypsinised (by addition of 300µl 1x Trypsin +0.25% EDTA) and fixed (by adding 300µl of 1x PBS +2% paraformaldehyde (PFA)).

When using suspension cell lines, cells were diluted to 10^5/ml. 24 hours later, cells were pelleted, supernatant discarded, and replaced with half the volume of media. 250µl was seeded per well of a 24 well plate. Another 250µl media containing the appropriate concentration of virus and PB is added. 48-72 hours later, cells were fixed (by addition of 250µl of 1x PBS + 2% PFA).

In all experiments, an uninfected control well was included to aid analysis. Each virus was titrated on each cell line before experimentation; starting from neat supernatant, a 2- or 3-fold serial dilution of 10-12 points was carried out. For all subsequent experiments, at least 4 viral dilutions were carried out, with the aim of achieving between 1-30% infection as assessed by flow cytometry (see section 2.8) or achieving data within the linear region of the assay when measuring Luciferase activity (see section 2.9).

2.6.2   Infections for qPCR

For all infections for qPCR, infections for flow cytometry/Luciferase assay were carried out in parallel. Adherent cell lines were seeded in 6 well plates at $5x10^4$/ml and 2ml/well. 24 hours later, media was discarded and replaced with 250µl media containing the appropriate concentration of DNase I treated virus (see section 2.5.2) and 8µg/ml PB. To ensure that DNase treatment of virus was successful, virus that had been boiled for 10mins (which destroys all infectious virus but does not affect contaminating DNA) was also added to one well. Samples were collected at 6hrs post-infection when measuring early/late reverse transcripts, or at 18hrs post-infection when measuring 2-LTR circles, unless otherwise stated. Samples were collected by trypsinising, pelleting, discarding the supernatant, washing with 1x PBS, pelleting, discarding the supernatant, and freezing at -20°C. For qPCR protocol see section 2.11.

2.6.3 Drug assays

Unless otherwise stated, drugs were used as described in Table 7.

Table 7: Mammalian cell culture drugs

Drug	Concentration (unless otherwise specified)	Time of Addition	Supplier
Aphidicolin (ApC)	2µg/ml	At seeding (and replenished at infection)	Calbiochem
Cyclosporine A (CsA)	5µM	At infection	Sandoz
PF-3450074 (PF74)	≤20µM		James.L
SmBz	10µM		Selwood. D
BI-1	as specified		Oxenford. S
SAL compounds	as specified		Warner. J

2.7 Protein depletion and expression

2.7.1 RNAi target and shRNA oligonucleotide design

19mer RNAi target sequences were designed using RNAi Target Sequence Selector (http://bioinfo.clontech.com/rnaidesigner/sirnaSequenceDesignInit.do, Clontech), and checked using the Basic Local Alignment Search Tool (BLAST, NCBI) for complementarity to off-target sequences. shRNA sequences were designed using shRNA Sequence Designer (http://bioinfo.clontech.com/rnai designer/oligoDesigner.do, Clontech) to include the hairpin loop TTCAAGAGA, an MluI site, and – as overhangs - a 5' BamHI site and a 3' EcoRI site. Typically, 3 target sequences were designed and tested for each target. RNAi targets used in this study are shown in Table 8.

2.7.2 Oligonucleotide annealing and pSIREN cloning

Oligonucleotides (Sigma) were reconstituted to 1µg/µl. 10µl of each primer were added to 2.5µl of 2M NaCl, heated to 98°C for 5 mins, and slowly cooled (approximately 6°C/min) to 4°C using a GS4 Multi Block Thermal Cycler (G-Storm). To this, 350µl dH$_2$O, 40µl 3M NaAcetate, and 1.1ml 100% ethanol, was added. This was frozen at -80°C for approximately 18hrs. DNA was pelleted by centrifugation at 13,000 rpm and 4°C for 45mins. Ethanol was removed as much as possible and left to air dry, before resuspending in 50µl dH$_2$O.

Table 8: RNAi target sequences and shRNA sequences

Target	RNAi target sequence	Oligonucleotide sequence (lower case are overhangs)	Notes
Control – 1	UCGGCGCCAGUCUAAUUAUA	gatccGTTATAGGCTCGCAAAAGGTTCAAGAGACCTTTTGCGAGCCTATAAACTTTTTTTACGCGTg aattcACGCGTAAAAAAGTTTATAGGCTCGCAAAAGGTCTCTTGAACCTTTTGCGAGCCTATAACg	(242)
Control – 2	GUUAUAGGCUCGCAAAAGG		
CPSF6 – 1	GCCAGAAAGACCGAGAUUAC	gatccGCCAGAAAGACCGAGATTACATTCAAGAGATGTAATCTCGGTCTTTCTGGTTTTTTACGCGTg aattcACGCGTAAAAAACCAGAAAGACCGAGATTACATCTCTTGAATCTCGGTCTTTCTGGCg	Strong depletion
CPSF6 – 2	GCGAAGAGUUCAACCAGGA	gatccGCGAAGAGTTCAACCAGGAATTCAAGAGATTCCTGGTTGAACTCTTCGCTTTTTTACGCGTg aattcACGCGTAAAAAACGAAGAGTTCAACCAGGAATCTCTTGAATCTCCTGGTTGAACTCTTCGCg	Strong depletion
CPSF6 – 3	GGUGGACAACAGAUGAAGA	gatccGGTGGACAACAGATGAAGATTCAAGAGATCTTCATCTGTTGTCCACCTTTTTTTACGCGTg aattcACGCGTAAAAAAGGTGGACAACAGATGAAGATCTCTTGAATCTTCATCTGTTGTCCACCg	Weak depletion
CPSF5 – 1	UGCGCAGUGGGAGAAGUAAA	gatccGTGCGCAGTGGGAGAAGTAAATTCAAGAGATTTACTTCTCCCACTGCGCACTTTTTTTACGCGTg aattcACGCGTAAAAAAGTGCGCAGTGGGAGAAGTAAATCTCTTGAATTTACTTCTCCCACTGCGCACg	Highly toxic
CPSF5 – 2	UGGGCACAUACUCCAAGAU	gatccGTGGGCACATACTCCAAGATTTCAAGAGAAATCTTGGAGTATGTGCCCACTTTTTTACGCGTg aattcACGCGTAAAAAATGGGCACATACTCCAAGATTTCTTGAAATCTTGGAGTATGTGCCCACg	Weak depletion
CPSF5 – 3	UGCAGAGAACUCCCAGUAU	gatccGTGCAGAGAACTCCCAGTATTTCAAGAGATATACTGGGAGTTCTCTGCATTTTTTTACGCGTg aattcACGCGTAAAAAATGCAGAGAACTCCCAGTATCTCTTGAAATACTGGGAGTTCTCTGCACg	Strong depletion
Nup153 - 1	CAAUUCGUCUCAAGCAUUA	gatccGCAATTCGTCTCAAGCATTATTCAAGAGATAATGCTTGAGACGAATTGTTTTTTTTACGCGTg aattcACGCGTAAAAAAACAATTCGTCTCAAGCATTATCTCTTGAATAATGCTTGAGACGAATTGCg	Strong depletion
Nup85 - 1	GAGCCAACAGUCACUUUGA	gatccGAGCCAACAGTCACTTTGATTCAAGAGATCAAAGTGACTGTTGGCTCTTTTTTTACGCGTg aattcACGCGTAAAAAAGAGCCAACAGTCACTTTGATCTCTTGAATCAAAGTGACTGTTGGCTCg	Highly toxic
Nup85 - 2	CCUCCAGAGAAUUGACGAA	gatccGCCTCCAGAGAATTGACGAATTCAAGAGATTCGTCAATTCTCTGGAGGCTTTTTTTACGCGTg aattcACGCGTAAAAAAGCCTCCAGAGAATTGACGAATCTCTTGAATTCGTCAATTCTCTGGAGGCg	Weak depletion
Nup85 - 3	CUGAGAACUGCUUCAAUGU	gatccGCTGAGAACTGCTTCAATGTTTCAAGAGACATTGAAGCAGTTCTCAGTTTTTTTACGCGTg aattcACGCGTAAAAAACTGAGAACTGCTTCAATGTTCTCTTGAAACATTGAAGCAGTTCTCAGCg	Strong depletion
NUP155 – 1	UCCAGAGAUCAGUUCCAUC	gatccGTCCAGAGATCAGTTCCATCTTCAAGAGAGATGGAACTGATCTCTGGATTTTTTTACGCGTg aattcACGCGTAAAAAATCCAGAGATCAGTTCCATCTCTCTTGAAGATGGAACTGATCTCTGGACg	Strong depletion
NUP155 – 2	CAGGCAUCUUUCAACCUCA	gatccGCAGGCATCTTTCAACCTCATTCAAGAGATGAGGTTGAAAGATGCCTGCTTTTTTTACGCGTg aattcACGCGTAAAAAACAGGCATCTTTCAACCTCATCTCTTGAATGAGGTTGAAAGATGCCTGCg	Highly toxic
Nup155 – 3	UCCAGAGAUCAGUUCCAUC	gatccGCTAGAGTCAGTGCTACAAATTCAAGAGATTGTAGCACTGACTCTAGCTTTTTTTACGCGTg aattcACGCGTAAAAAAGCTAGAGTCAGTGCTACAATCTCTTGAATTGTAGCACTGACTCTAGCg	Highly toxic
Nup358	GCGAAGUGAUGAUAUGUUU		(79)
TNPO3	UCGGCGCCACAGAAAUUAUA		(242)

Annealed oligos were subsequently cloned into pSIREN RetroQ by digestion with BamHI and EcoRI (see section 2.3). Colonies were screened by analytical digestion with MluI, which is found within the oligonucleotide but not elsewhere in pSIREN. Positive ligations were confirmed by sequencing (Beckman).

2.7.3 Transduction of mammalian cell lines with shRNA/transgene-expressing vectors

Both RNAi Ready pSIREN RetroQ- and pEXN-based retroviral expression vectors were produced by 3-plasmid transfection as described in section 2.5.1 and Table 6. In all experiments, a control vector was also used. In the case of pSIREN RetroQ-based vectors, a non-targeting shRNA was used. In the case of pEXN-based vectors, the parental plasmid pEXN (which encodes no protein) was used. The following protocol applies to both depletion and expression experiments.

When using adherent cell lines, cells were seeded in 6 well plates at 5×10^4/ml and 2ml/well. 24 hours later, media was discarded and replaced with 1ml fresh media and 1ml vector and PB. 48hrs later, cells were thoroughly trypsinised (300μl), counted, and reseeded for infection as described in section 2.6.1. When using suspension cell lines, cells were diluted to 10^5/ml. 24 hours later, cells were pelleted, supernatant discarded, and replaced with in half the volume of media. 1ml was seeded per well of a 6 well plate. 1ml vector was added to each well. If toxicity was problematic, the amount of vector was reduced down to 500μl or 250μl/well.

Initially, all depletion experiments cells were carried out without selection as follows; 48hrs post-transduction, cells were reseeded for infection as described in section 2.6.1. A transient depletion such as this was preferable as it reduced the duration of the experiment and thus the likelihood of toxicity. However, if this protocol led to inadequate depletion, cells were selected with puromycin ((Merck) used at 1μg/ml) at 48 hours post-transduction for 48-72hrs (or until

an untransduced control well was completely killed) before reseeding and infection. In all protein expression experiments, cells were selected with G418 (Geneticin G-418 sulphate (Invitrogen) used at 500μg/ml) was added at 48hrs post-transduction for 5-10days (or until an untransduced control well was completely killed).

At the time of infection, samples were also collected for Western blotting (see section 2.13) or SYBR Green qPCR (see section 2.12) in order to assess the extent of depletion or expression.

HeLa cells clones stably expressing the shRNAs against Control A, Nup358 and TNPO3, were provided by T.Schaller (79), and were maintained under puromycin selection.

Where simultaneous depletion experiments were performed, the less toxic depletion was carried out first (this may or may not have included selection), and at least 72hrs were left between the first and second transductions. Where simultaneous expression and depletion experiments were performed, a stably expressing cell line was created first, and the depletion performed second.

2.8 Flow cytometry

The proportion of infected, and thus GFP expressing, cells was determined by flow cytometry using the BD Accuri C6 (BD Biosciences). Live cells were gated by forward scatter height (FSC-H) and side scatter height (SSC-H), which describe size and granularity respectively. 10,000 alive cells were measured for each sample. An uninfected negative control sample was used to design a gate for GFP-positive cells, as measured by FLH-1, which describes fluorescence of 515-545nm wavelengths.

Only values between 1% and 30% were used for analysis. Below or above these levels we did not observe a linear relationship between viral input and infectivity, suggesting that one GFP-positive cell did not represent one

infectious viral particle or infectious unit (IU), which we sought to measure. The Poisson distribution tells us that, above 35% infectivity, there is a dramatic increase in the chances of one GFP positive cell being infected by two viruses and thus representing more than one IU. Values of percentage infectivity were normalised by viral input as follows;

$$\frac{\text{percentage of cells GFP}^+ \text{ x number of cells in well}}{\text{ml virus in well}} = \text{IU/ml}$$

The viral input (for HIV-1 viruses only) could also be normalised by ng of RT per ml of virus as calculated by RT-ELISA (described in section 2.5.3), resulting in a value measured in IU/ngRT. For each condition tested, viral titres were calculated from three different viral inputs. The mean and standard deviation of these triplicate values were then determined.

2.9 Luciferase assay

Cells that had been infected with HIV-1(Luc) for 48 hours were lysed by adding 100µl of Permeabilisation Buffer (Biotium) per well and freezing for at least 30mins. This was thawed and mixed vigorously by pipetting. 20µl was transferred into a well of a 96-well GloMax Luminometer Light Plate (Promega). 30µl of SteadyGlo Luciferase Assay Substrate (diluted as per instructions in SteadyGlo Luciferase Assay Buffer) (both Promega) was then added to each well. The number of relative light units (RLU) were read 3 times using a GloMax 96 Microplate Luminometer (Promega) and an average was taken.

An uninfected negative control was used to measure any background light and this value was subtracted from all other measurements. As discussed in section 2.8, a titration of virus was used to determine the amount of viral supernatant that resulted in a linear relationship between viral input and RLU output. For each sample, at least three titration points within the linear range were used to calculate the average RLU/ml virus.

2.10 Nucleic acid extraction and reverse transcription

2.10.1 DNA extraction

DNA was extracted from mammalian cells using QIAamp DNA Blood Mini Kit (Qiagen). DNA concentrations of 25-200ng/µl, as measured by nanodrop, were obtained. DNA was stored at -20°C.

2.10.2 RNA Extraction

RNA was extracted from mammalian cells using the RNeasy Mini Kit (Qiagen), including the optional 'On-Column DNase Digestion' step. RNA was stored at -80°C.

2.10.3 cDNA synthesis

1µg of RNA, as measured by nanodrop, was reverse transcribed into cDNA using Omniscript RT Kit (Qiagen). cDNA was stored at -20°C.

2.11 Taqman quantitative PCR (qPCR)

2.11.1 Principles of fluorescent reporter probe qPCR

Taqman reporter probes comprise a target-DNA binding sequence with a fluorophore at one end and a fluorescence quencher at the other. In this study, the fluorophore used was FAM (6-carboxyfluorescein) and the quencher used was TAMRA (tetramethylrhodamine). When unbound or when bound to ssDNA, the quencher is in close proximity to the reporter and so fluorescence is quenched. qPCR probes are designed to hybridise to the region of DNA amplified by the primers. Therefore, the annealing step of the PCR causes both primers and the probe to anneal to target ssDNA. During extension, *Taq* polymerase extends 3'-5' from primers and simultaneously uses its 5'-3' exonuclease activity to degrade the probe. Degradation of the probe results in the fluorophore being released such that it is no longer in proximity to the quencher. Therefore, the fluorescence detected by a qPCR thermal cycler is directly proportional to the amount of target DNA presence in the reaction.

2.11.2 TaqMan primers and probes

The DNA targets, primer and FAM/TAMRA-probe sequences (Sigma), and plasmid standards used in this study are detailed in

Measurement	DNA Target	Primers and probes	Standard
HIV-1(GFP) late reverse transcripts	eGFP	GT139 FWD: CAACAGCCACAACGTCTATATCAT GT140 REV: ATGTTGTGGCGGATCTTGAAG GFP probe: FAM-CCGACAAGCAGAAGAACGGCATCAA-TAMRA	pCNCG
HIV-1 early reverse transcripts	LTR 'R'	OHC64 FWD: TAACTAGGGAACCCACTGC OHC65 REV: GCTAGAGATTTTCCACACTG OHC66 probe: FAM-ACACAACAGACGGGCACACACTA-TAMRA	pCSGW
HIV-1 2-LTR circles	LTR-LTR junction	2-LTR LA FWD: AACTAGAGATCCCTCAGACCCTTTT 2-LTR LA REV: CTTGTCTTCGTTGGGAGTGAATT 2-LTR junction probe: FAM-TTCCAGTACTGCTAGAGATTTTCCACACT-TAMRA	p2LTR-LA

.

Table 9: TaqMan qPCR primers, probes, and standards

Measurement	DNA Target	Primers and probes	Standard
HIV-1(GFP) late reverse transcripts	eGFP	GT139 FWD: CAACAGCCACAACGTCTATATCAT GT140 REV: ATGTTGTGGCGGATCTTGAAG GFP probe: FAM-CCGACAAGCAGAAGAACGGCATCAA-TAMRA	pCNCG
HIV-1 early reverse transcripts	LTR 'R'	OHC64 FWD: TAACTAGGGAACCCACTGC OHC65 REV: GCTAGAGATTTTCCACACTG OHC66 probe: FAM-ACACAACAGACGGGCACACACTA-TAMRA	pCSGW
HIV-1 2-LTR circles	LTR-LTR junction	2-LTR LA FWD: AACTAGAGATCCCTCAGACCCTTTT 2-LTR LA REV: CTTGTCTTCGTTGGGAGTGAATT 2-LTR junction probe: FAM-TTCCAGTACTGCTAGAGATTTTCCACACT-TAMRA	p2LTR-LA

2.11.3 Taqman qPCR setup

2x TaqMan Gene Expression Master Mix was used in these experiments and contains the following components; AmpliTaq Gold DNA Polymerase Ultra Pure, Uracil-DNA glycosylase (not required in these experiments), dNTPs and dUTP, ROX passive reference, and optimized buffer components. A master mix was made as detailed in Table 10 and 15μl transferred to each well of a MicroAmp Fast Optical 96-Well Reaction Plate with Barcode 0.1 mL (Applied Biosystems).

5µl of DNA (either DNA extracted from mammalian cells (see section 2.10.1) or plasmid DNA standards) was then added to each well and mixed gently by pipetting. 6 plasmid standards were used in each experiment at concentrations of 10^1, 10^2, 10^3, 10^4, 10^5, and 10^6 copies/µl in 10µg/ml sheared salmon sperm carrier DNA (Thermo Fischer). Each DNA sample/standard was tested in triplicate. A negative control well in which DNA was replaced by dH_2O was included in order to ensure there was no contamination of the master mix. PCR plates were sealed with MicroAmp Optical Adhesive Film (Applied Biosystems) and briefly centrifuged to collect liquid and remove any bubbles.

Table 10: TaqMan qPCR reagents

Component/ Stock Concentration	Volume (µl)
2x TaqMan Gene Expression Master Mix	10
FAM/TAMRA Probe 7.5µM	0.5
Primer 1 7.5µM	1
Primer 2 7.5µM	1
dH_2O	2.5
Master Mix Total	**15µl**
DNA sample (20-150ng/ul) or plasmid standard (10^1-10^6 copies/µl)	5
Total	**20µl**

2.11.4 TaqMan qPCR cycling parameters

qPCRs were carried out using a 7500 Fast Real-time PCR System (Applied Biosystems) as detailed in table Table 11. The system was instructed to detect FAM as the reporter, TAMRA as the quencher, and ROX (included in the TaqMan master mix) as a reference dye.

Table 11: TaqMan qPCR cycling parameters

Step	Temp (°C)	Time		No. of Cycles	
		GFP	2LTR	GFP	2LTR
Polymerase Activation	95	10mins		1	
Denaturing	95	15secs		40	50
Annealing/extension	60	60secs	90secs		

2.11.5 TaqMan qPCR analysis

To rule out the possibility of plasmid DNA contamination of the virus and the extracted DNA it was ensured that all boiled and negative controls, respectively, had undetectable levels of target DNA. Like all PCRs, qPCR has an exponential phase in which each cycle results in the number of copies being doubled, before it plateaus. It was ensured that all Ct FAM (cycle threshold FAM, the cycle at which FAM was first detectable) values were taken from the exponential phase of the reaction (visualised by plotting DNA copies against cycle number). A standard curve was created by plotting the original target DNA copy number (which is known for the standards) against Ct FAM value. The equation of this curve was then used to convert the Ct FAM value for each well into the original number of target copies in that well. This copy number was normalised to the amount of DNA put into each well and also to the amount of virus used (either measured in ml or ngRT). Each sample was measured by qPCR in triplicate and an average taken. Each experimental condition was carried out using at least 3 different viral doses, from which an average and standard deviation could be calculated.

2.12 SYBR Green qPCR

2.12.1 Principle of dsDNA-binding dyes as qPCR reporters

SYBR Green I Dye binds non-specifically to dsDNA and fluoresces when it does so. The level of fluorescence detected by a qPCR thermal cycler is therefore directly proportional to the amount of DNA presence in the reaction.

2.12.2 SYBR Green qPCR primers

The forward primer for detection of CPSF5 was located in exon 1 and the reverse primer in exon 2. Leaving >500bp of intron between primers ensured that only cDNA, and not any contaminating gDNA, was measured. The housekeeping gene β-2-microglobulin used as a control (provided by Helen Rowe). SYBR Green qPCR primers used in this study are provided in Table 12.

Table 12: SYBR Green qPCR primers

Target	Primers
CPSF5	JCR99: 5'- CTCACCCTGGAGCGCACCATC-3'
	JCR100: 5'- GGGTAGCCGGTGCTCATGTAC-3'
β-2-microglobulin	HuB2MF: 5'-TGATCGCGCTACTCTCTCTTT-3'
	HuB2MR: 5'-TCTGCTGGATGACGTGAGTAAAC-3'

2.12.3 SYBR Green PCR setup

2x SYBR Green PCR Master Mix (Applied Biosystems) was used in these experiments and contains the following components; SYBR Green I Dye, AmpliTaq Gold DNA Polymerase, dNTPs, ROX passive reference dye, and optimised buffer components. For each primer pair, a master mix was made as detailed in Table 13 and 20μl transferred to each well of a MicroAmp Fast Optical 96-Well Reaction Plate with Barcode 0.1 mL (Applied Biosystems). 5μl of cDNA (500ng total) was then added to each well and mixed gently by pipetting. Each cDNA sample was tested in triplicate. A negative control well in which cDNA was replaced by dH_2O was included in order to ensure there was no contamination of the master mix. PCR plates were sealed with MicroAmp Optical Adhesive Film (Applied Biosystems) and briefly centrifuged to collect liquid and remove any bubbles.

Table 13: SYBR Green qPCR reagents

Component/ Stock Concentration	Volume (μl)
2x SYBR Green PCR Master Mix	12.5
Primer 1 3μM	2
Primer 2 3μM	2
dH_2O	3.5
Master Mix Total	**20μl**
DNA sample (100ng/ul)	5
Total	**25μl**

2.12.4 SYBR Green PCR cycling parameters

qPCRs were carried out using a 7500 Fast Real-time PCR System (Applied Biosystems) as detailed in Table 14. The system was instructed to detect SYBR as the reporter and ROX as a reference dye.

Table 14: SYBR Green qPCR cycling parameters

Step	Temp (°C)	Time	No. of Cycles
Polymerase activation	95	10mins	1
Denaturing	95	15secs	40
Annealing/extension	60	60secs	
Melt Curve	95	15secs	1
	60	60secs	1
	60 to 95 +0.3°C/sec	Approx. 2mins total	1
	95	30secs	1
	60	15secs	1

Due to the fact that SYBR Green I Dye binds non-specifically to dsDNA, it was necessary to test that these primers only bound to one specific target. This was done using a melt-curve analysis, also known as a dissociation analysis, which identifies the temperature at which 50% of dsDNA dissociates (which causes the SYBR Green level to drop as it does not bind ssDNA). If two different temperature peaks appear, this suggests that two different regions of DNA (with different T_ms) have been amplified. This was not found to be a problem with the primers described above.

2.13 Western blotting

2.13.1 Sample preparation

Laemmli Buffer (4% SDS, 20% glyercol, 0.004% bromophenol blue, 0.125M Tris HCl, pH6.8) was supplemented with 5% freshly added β-mercaptoethanol

(Sigma) just before addition to cells. $2x10^5$ cells were pelleted and lysed in 100µl Laemmli buffer+β-ME. Unless otherwise stated, samples were kept on ice. Samples were sonicated for 10secs and then boiled for 10mins. Samples were briefly centrifuged to collect liquid. Samples were stored at -80°C.

2.13.2 Sodium dodecyl sulphate (SDS) polyacrylamide gel preparation

SDS polyacrylamide gels were prepared as detailed in Table 15. Tetramethylethylenediamine (TEMED) and ammonium persulfate (APS) were added immediately before pouring the gels. Usually, 10% resolving gels were used. However, when blotting for proteins larger than 250kDa (e.g. Nup358) 7% resolving gels were used. The Mini-PROTEAN Tetra System (BioRad) was used for casting and running all gels. Resolving gels were poured first and topped with a layer of isopropanol in order to ensure an even interface with the stacking gel. Once set, isopropanol was removed, rinsed with dH_2O, and any remaining liquid dried with blotting paper. The resolving gel was then overlaid with stacking gel, into which a 10- or 15-well comb was inserted.

Table 15: SDS polyacrylamide gel recipe

Component	Supplier	10% resolving	7% resolving	6% stacking
Acrylamide	Flowgen	10ml	7ml	2ml
Bis-acrylamide	Bioscience	4ml	2.8ml	800µl
SDS 20%		150µl	150µl	50µl
1M Tris pH8.8 (running)		11.25ml	11.25ml	
0.5M Tris pH6.8 (stacking)				2.5ml
dH_2O		4.5ml	8.7ml	4.6ml
TEMED	Sigma	10µl	10µl	10µl
APS 20%	BioRad	150µl	150µl	50µl
Total		**30ml**	**30ml**	**10ml**

2.13.3 Polyacrylamide gel electrophoresis (PAGE)

The Mini-PROTEAN Tetra System (BioRad) was used to run all Western blots. Samples were run submerged in 1x running buffer (10x; 144g glycine, 30g Tris base, 10g SDS, made to 1L with dH_2O). 10-20µl of sample was loaded per well,

alongside 10μl PageRuler Plus Prestained Protein Ladder (Fermentas). Samples were run at 150V for 1-2hrs, or until the desired ladder separation has occurred.

The Mini Trans-Blot Electrophoretic Transfer Cell System (BioRad) was used to wet-transfer all Western blots. Samples were transferred submerged in 1x transfer buffer (10x; 144g glycine, 30g Tris base, made to 1L with dH$_2$O), onto either Hybond polyvinylidene difluoride (PVDF) membrane (GE Healthcare) or Immobilon PVDF membrane (Millipore). Membranes were activated in methanol for 1min before layering onto gels. This gel-membrane sandwich was surrounded by transfer buffer-soaked Whatman blotting paper (Sigma). Transfers were carried out at 4°C, either at 100V for 2hrs or at 20V for 18hrs.

After transfer, membranes were blocked for non-specific antibody binding by incubation in 1x TBS (Tris buffered saline 10x; 120.5g Tris, 400g NaCl, made up to 1L with dH$_2$O, pH7.6) +0.1% Tween 20 (BioRad) +5% powdered milk, on a rocking platform at room temperature for at least 1hr.

2.13.4 Blotting

All antibodies were diluted in TBS-T (Tris buffered saline +0.1% Tween), and incubated with membranes on a rocking platform at room temperature for 1hr. The antibodies used in this study and the conditions in which they were used are detailed in Table 16. After incubation with the primary antibody, membranes were washed in TBS-T, 4 times for 20mins each. Membranes were then incubated with the appropriate horse radish peroxidase (HRP)-conjugated secondary antibody, again for on a rocking platform at room temperature for 1hr. Membranes were washed as before.

Table 16: Antibodies

Antibody type	Target	Expected weight (kDa)	Antibody ID and supplier	Organism raised in	Dilution used at
Primary (unconjugated)	β-Actin	42	Ab6276 Abcam	Mouse	1/20000
	CPSF6	68	Ab99347 Abcam	Rabbit	1/2000
	CypA	18	SA296 Biomol	Rabbit	1/5000

	Nup153	153	Ab24700 Abcam	Mouse	1/2000
	Nup358	358	C288 Catavas.E	Rabbit	1/2000
	Nup155	155	Ab73292 Abcam	Rabbit	1/2000
	Nup85	85	45-102 ProSci	Goat	1/1000
	TNPO3	95	Ab54353 Abcam	Mouse	1/100
Secondary (HRP conjugated)	Goat Abs		RPN4301 GE	Rabbit	1/5000
	Mouse Abs		NXA931 GE	Sheep	1/5000
	Rabbit Abs		NX934V GE	Donkey	1/5000
Primary (HRP conjugated)	HA (HA tags)		3F10-HRP Roche		

Membranes were treated with Amersham ECL Select or Prime Western Blotting Detection Reagent (GE Healthcare) or RapidStep ECL Reagent (Calbiochem). The HRP bound to the secondary antibody cleaves these chemiluminescent substrates, resulting in luminescence that is detected by Amersham Hyperfilm ECL Film (VWR), a process known as autoradiography.

In order to blot for a second protein on the same membrane, bound antibodies were removed using a pH2.2 glycine stripping buffer (15g glyine, 1g SDS, 10ml Tween 20, made to 1L with dH$_2$O, pH2.2, kept at 4°C) for 30mins on a rocking platform. Membranes were then blocked and probed with antibodies as described above.

2.14 SureSelect^{XT2} integration site analysis

To avoid the bias introduced by other methods (such as Alu-Gag PCR and MluI digestion), randomly sheared DNA from infected cells integration sites was enriched for provirus-containing DNA using SureSelect^{XT2}, which uses beads coated in baits designed against HIV-1 LTRs.

2.14.1 Infections for SureSelect^{XT2} integration site analysis

6x10^5 HeLa cells were seeded into a 10cm dish. 24 hours later (approximately 1.2x10^6 cells present), cells were infected with DNase treated viral vector (see sections 2.5.2 and 2.6.1) in a total of 6ml with a multiplicity of infection (MOI) of ~0.8. Cells were maintained for 7 days, which ensured that cytoplasmic,

unintegrated reverse transcripts were mostly degraded and thus their LTRs would not saturate the SureSelect[XT2] baits. Cells were never split more than 1:4 every 2 days, which ensured that integration site diversity was not lost over the 7 days. On day 7, cells were counted and samples of 10^6 collected. DNA was extracted as described in section 2.10.1. DNA concentration was measured using the Quant-iT dsDNA BR Assay Kit with a Qubit 2.0 Fluorometer (both Life Technologies).

2.14.2 DNA shearing

DNA was sheared as described in the SureSelect[XT2] Target Enrichment System (Illumina) protocol, with the changes described below. In part 2 step 1, 3µg of DNA was made to 130µl using 1x Low TE Buffer (Applied Biosystems) in a DNA LoBind Tube, 1.5ml PCR Clean (Eppendorf). An S220 Focused Ultrasonicator Sample Preparation System (Covaris) was used to shear the DNA to a target fragment size of 300bp.

2.14.3 SureSelect[XT2] target enrichment system kit

The SureSelect[XT2] Target Enrichment System for Illumina (Agilent) consisted of SureSelect[XT2] Reagent Kit for MiSeq Platform for 16 Samples and SureSelect[XT2] Capture Library 5190-4846. The capture library was custom designed to have baits targeting the LTRs of HIV-1 and HIV-2. Baits were 120nt each, shifted by 1nt each, and covered the whole 370nt SIN LTR (resulting in 250 different baits), which in pCSGW is as follows;

```
GATCTCCCTTTGGGCCGCCTCCCCGCATCGATACCGTCGACCTCGAGGGAATTAATT
CGAGCTCGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACT
TTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGAT
CTGCTTTTTGCTTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTC
TCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTT
CAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCC
TTTTAGTCAGTGTGGAAAATCTCTAGCAG
```

The SureSelect[XT2] protocol was followed with the following adjustments. Between part 2 steps 1 and 2, DNA was cleaned up using Agencourt AMPure XP Beads (Beckman Coulter) and eluted in 50µl. In all bioanalysis steps, either the

DNA 1000 Kit or the High Sensitive DNA Kit was used with a 2100 Bioanalyzer (all Agilent). In part 2 step 8, 6 cycles of PCR were performed. In part 3 step 1, 8 indexed gDNA samples were pooled for hybridisation. In part 3 step 3, Dynabeads MyOne Streptavidin T1 (Life Technologies) magnetic beads were used. In part 4 step 1, 14 cycles of PCR were carried out. All samples were sequenced by UCL genomics using the Illumina Miseq platform.

2.14.4 Integration site analysis

All computational analysis was carried out by Alessandro Riccombeni as follows; for all libraries, FastQC (http://www.bioinformatics.babraham.ac.uk/ projects/fastqc/) was used to perform quality control. A custom Python script was used to remove potential PCR duplicates from each library; in case of multiple fastq records with the same sequence, only one was kept. Low quality bases were trimmed using seqtk (https://github.com/lh3/seqtk). Reads were locally mapped on the proviral construct sequence using Bowtie2 (PMID: 22388286) with --very-sensitive-local settings and maximum one reported mapping per read. Reads which completely mapped on the construct sequence were discarded. Partially mapping reads were split, and the soft clipped parts were mapped using tophat2 (PMID: 23618408) with very sensitive settings and maximum one mismatch on version hs19 of the human genome. Custom Python scripts were used for the downstream analysis. Unmapped reads and reads mapping to multiple positions were discarded. The mapping of each hard clipped read was matched to the corresponding whole reads' partial mapping against the construct sequence, obtaining the coordinates of viral integration sites. Controls were prepared for each library by taking a random 10% subset of all reads mapping completely on the human sequence and considering the first mapping position as a control site. To estimate gene and CpG densities around sites, a compressed version of the RefSeq annotation of hs19 was used. RefSeq gene and CpG islands annotations were downloaded from UCSC (PMID: 22908213). For each strand, overlapping features were merged and features whose coordinates were nested or identical to another feature's were

discarded. Density scores and the ratios of sites within genes were clustered using Pvclust (PMID: 16595560).

2.15 Statistical analysis

Having calculated mean and standard deviation values for our samples, we sought to test whether the difference between two samples was statistically significant. To do this we utilised the unpaired/independent two-sample Student's t-test, which tests the null hypothesis that samples 1 and 2 are equal. It assumes that variation within samples follows the normal distribution. The 't-statistic' (t) is calculated using the mean values of samples 1 and 2 ($\bar{X}_1$ and $\bar{X}_2$), the standard deviation of samples 1 and 2 (s_1 and s_2), and the number of data points within samples 1 and 2 (n_1 and n_2) as follows;

$$t = |\overline{X_1} - \overline{X_2}| \div \sqrt{A \times B}$$

where

$$A = (n_1 + n_2) \div n_1 n_2$$

and

$$B = \frac{[(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2{}^2]}{[n_1 + n_2 - 2]}$$

The t-statistic is then compared to the t-distribution (using a t-distribution probability table) to find a p-value (calculated probability). The more the t-statistic diverges from the t-distribution, the lower the p-value. p-values are conventionally considered 'statistically significant' when $p \leq 0.05$, which indicates that the probability of the null hypothesis being true is ≤5%. In other words, there is ≥95% certainty that the null hypothesis is false and that the two samples differ. In this study * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$, and *** indicates $p \leq 0.001$.

3 Chapter 3. The role of CPSF6 in HIV-1 nuclear entry

3.1 Introduction

As detailed in section 1.5.3, Lee *et al.* (259) hypothesised that endogenous CPSF6 is a restriction factor of HIV-1. They showed that exogenous expression of hCPSF6[72] inhibited the virus ~2-fold and depletion of endogenous CPSF6 increased HIV-1 infectivity ~1.2-fold. They showed that artificially C-terminally truncating CPSF6 such that its classical NLS is removed (mCPSF6[72]-358) caused it to mislocalise to the cytoplasm and to potently inhibit the virus. They proposed that endogenous CPSF6, which is predominantly nuclear at steady state, might localise to the cytoplasm under certain conditions but did not speculate as to what these conditions might be. We consider the evidence supporting CPSF6 as an HIV-1 restriction factor to be very modest. However, we were compelled by the evidence that CPSF6 interacts – be it directly or indirectly – with HIV-1 CA and that this interaction is conserved in primate Lentiviruses. We were also extremely interested in the observation that the mCPSF6[72]-358 escape mutant N74D utilises a different set of nuclear entry cofactors than WT HIV-1 and that this results in different integration site targeting (79). This phenotype is reminiscent of another CA mutant, G89V, which prevents interaction with the cofactor CypA (discussed in section 1.5.2.2). Based on the conservation of CA-CPSF6 interaction, the similarities between N74D and G89V, and the finding that N74D is defective in some cell types (specifically MDMs) we sought to test an alternative hypothesis that CPSF6 is a cofactor of HIV-1 infection and that C-terminal truncations of CPSF6 are dominant negatives.

3.2 Results

3.2.1 HA-hCPSF6[68]-358 restricts HIV-1 prior to reverse transcription, but HA-hCPSF6[68] does not affect HIV-1 infectivity

Lee, et al. (259) reported that hCPSF6[72] expression inhibited HIV infectivity 2-fold and that hCPSF6[72]-358 inhibited infectivity over 20-fold. However, it has previously been reported that CPSF6[68], from which exon 6 has been spliced out, is vastly more abundant than CPSF6[72] in human cells (313, 316). In agreement with this, CPSF6 cDNA cloned from HeLa cells was repeatedly found to encode the spliced 68kDa form (cDNA from Bieniasz. P). Furthermore, western blots of CPSF6 in HeLa cells revealed only one band of ~68kDa (figures 17B/D). We were therefore interested to test whether overexpression of this isoform would impact HIV-1 infectivity. We ligated CPSF6[68] cDNA into the expression vector pEXN, which causes transgenes to become 5' HA-tagged as depicted in figure 17A. We transduced HeLa cells with MLV-based vectors carrying genomes derived from either pEXN-HA-hCPSF6[68] or empty pEXN as a negative control. We then selected for successfully transduced cells using G418 (pEXN encodes a neomycin resistance gene) until untransduced control cells were all dead (approximately 5 days). We confirmed overexpression of HA-hCPSF6[68] by western blot (figure 17B). We then infected these cells with serial dilutions of the single-round vector HIV-1(GFP) and determined the percentage of infected cells 48 hours later by flow cytometry for GFP-positive cells. We found that overexpression of HA-hCPSF6[68] had no significant effect upon HIV-1 infectivity (figure 17C).

We next tested whether C-terminal truncation of HA-hCPSF6[68] at residue 358 restricts HIV-1 in the same way that mCPSF6[72]-358 has been shown to (259). To do this, we introduced a stop codon after Asn358 creating HA-hCPSF6[68]-358 (cloning by Fletcher. A, construct depicted in figure 17A). As before, we transduced HeLa cells to express this protein and selected with G418. This protein was estimated to be 41kDa and was expressed at a similar level to endogenous CPSF6 as measured by western blot (figure 17D). We infected these cells with HIV-1(GFP) and then, in parallel, tested for late reverse
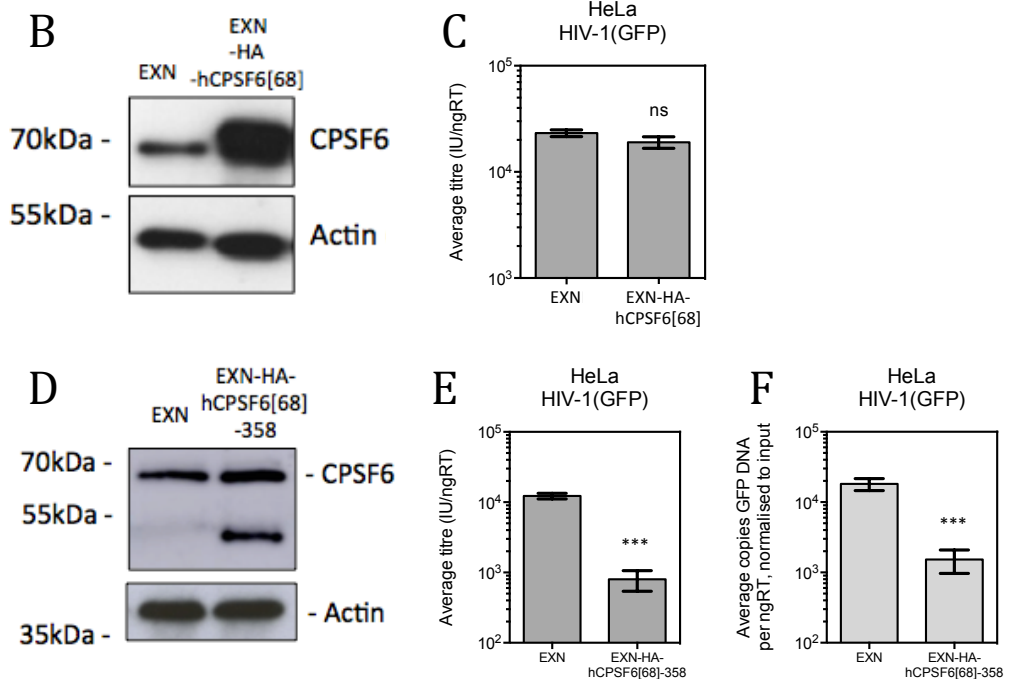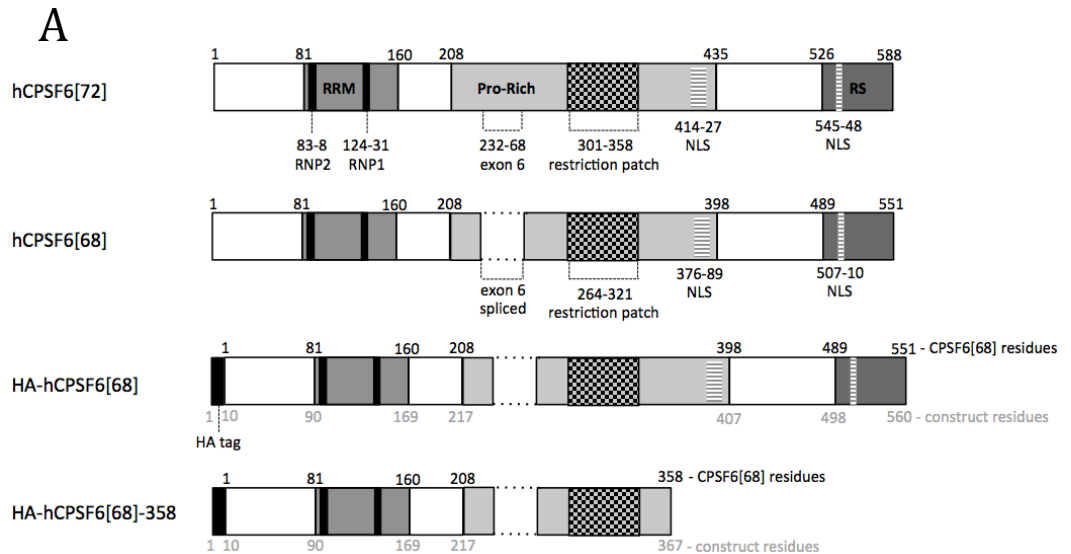
Figure 17: hCPSF6[68]-358, but not full-length hCPSF6[68], restricts HIV-1 prior to reverse transcription

(A) A schematic diagram of hCPSF6[72], hCPSF6[68], HA-hCPSF6[68], and HA-hCPSF6[68]-358. The checked box indicates the region of CPSF6 defined by Lee *et al.* as essential for CA-CPSF6 interaction **(B)** A western blot of lysates from HeLa cells stably expressing HA-hCPSF6[68] or empty vector using an anti-CPSF6 antibody. β-actin serves as a loading control. **(C)** HeLa cells from B were infected with a titration of HIV-1(GFP) and viral titre measured 48hrs p.i. by flow cytometry. **(D)** A western blot of lysates from HeLa cells stably expressing HA-hCPSF6[68]-358 or empty vector using an anti-CPSF6 antibody. β-actin serves as a loading control. **(E-F)** HeLa cells from (D) were infected with a titration of HIV-1(GFP). **(E)** Viral titre was measured 48hrs p.i. by flow cytometry. **(F)** Late reverse transcripts were measured 6hrs p.i. by qPCR. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean). ns means p>0.05 (not statistically significant, * means p≤0.05, ** means p≤0.01, *** means p≤0.001, as measured by Student's *t*-test.

transcripts by qPCR at 6hrs post-infection or for percentage infectivity at 48hrs post-infection. We found that HA-hCPSF6[68]-358 caused a 15–fold block to HIV-1 infection (figure 17E). Unlike mCPSF6[72]-358, which was reported to inhibit after reverse transcription (259), HA-hCPSF6[68]-358 inhibited reverse transcription by 12-fold (figure 17F). This suggests that neither exon 6 nor residues 322-358 of CPSF6[68] affect the ability of C-terminally truncated CPSF6 to restrict infection, but that one or both of these regions influences the stage of restriction. Lee *et al.* hypothesised that hCPSF6[72]-358 inhibited HIV-1 nuclear entry by preventing interaction between CA and a nuclear entry cofactor. However, the fact that two different C-terminal truncations of CPSF6 inhibit the virus at different stages of infection (before or after transcription) argues against this idea.

3.2.2   HA-hCPSF6[68]$_{180\text{-}321}$ is sufficient to inhibit HIV-1 nuclear entry, but not reverse transcription

Lee et al. showed that mCPSF6[72]-300 was unable to restrict HIV-1 or interact with HIV-1 CA-NC tubes, which demonstrated that CPSF6[68] residues 264-321 (which lie within the pro-rich domain) were necessary for interaction with the virus (259). We therefore sought to test whether this 58 residue region was sufficient to interact with and/or restrict the virus. Unfortunately, we were unable to express the peptide -hCPSF6[68]$_{264\text{-}321}$ even when fused to GFP as a carrier. We therefore tested a larger peptide, which comprised the entire pro-rich domain up to CPSF6[68] residue 321. The lower boundary of the pro-rich domain is described by Ruegsegger, et al. (313) as being residue 208 but the UniProt database (http://www.uniprot.org/uniprot/Q5ZL34) places this boundary as low as residue 180 (337). We therefore fused the peptide hCPSF6[68]$_{180\text{-}321}$ to the C-terminus of GFP before cloning it into the expression vector pEXN (construct shown in figure 18A). This protein had a predicted molecular weight of 44kDa and was expressed efficiently in HeLa cells as measured by western blot with an anti-HA antibody (figure 18B). We infected these cells with serial dilutions of HIV-1(Luc) and then tested either for early reverse transcripts 6 hours post-infection or for luciferase activity 48 hours
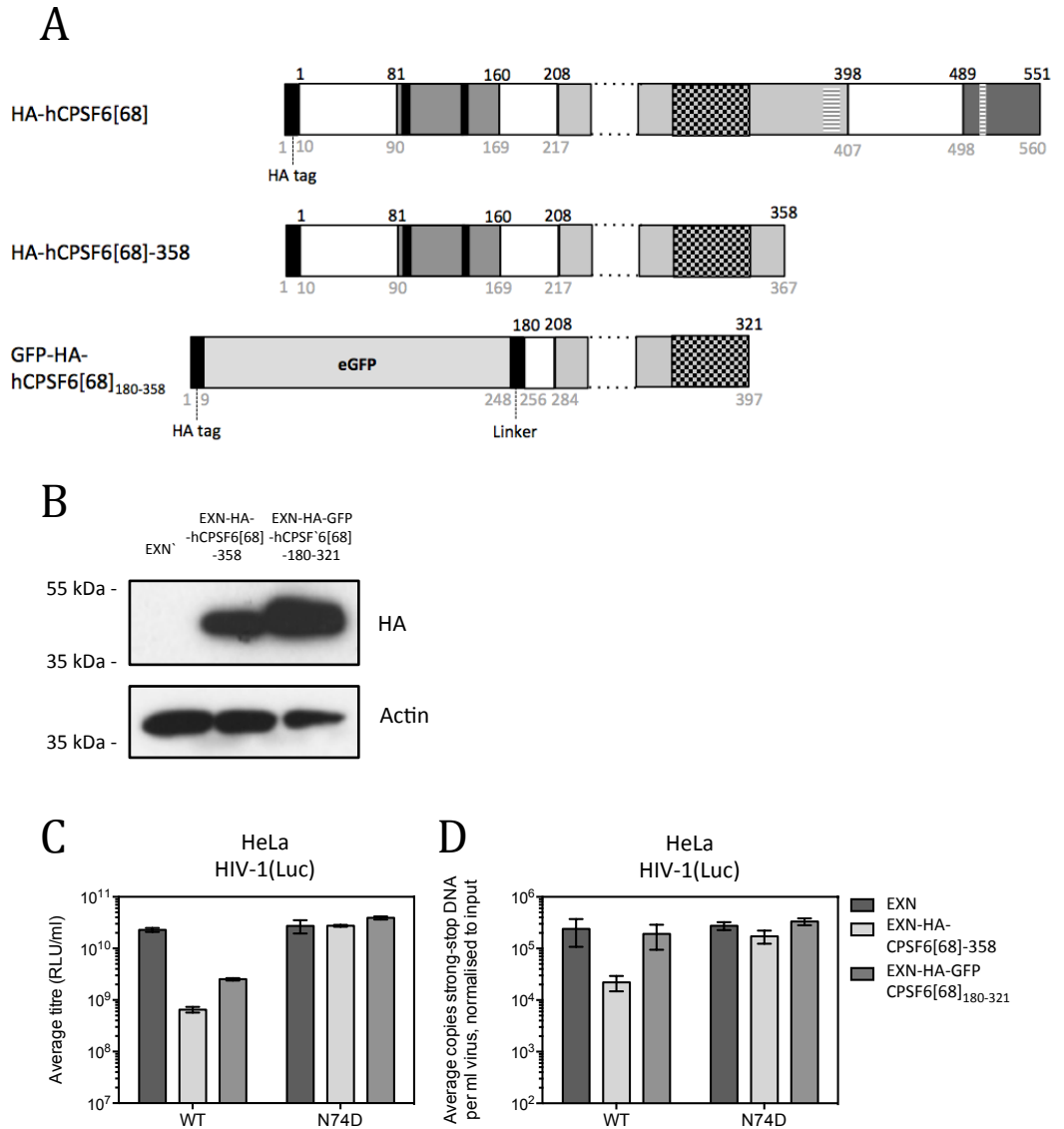
Figure 18: GFP-HA-hCPSF6[68]$_{180\text{-}321}$ restricts HIV-1 prior to reverse transcription

(A) A schematic diagram of HA-hCPSF6[68], HA-hCPSF6[68]-358, and GFP-HA-hCPSF6[68]$_{180\text{-}321}$. Shading as in figure 17A. **(B)** A western blot of lysates from HeLa cells stably expressing HA-hCPSF6[68]-358, GFP-HA-hCPSF6[68]$_{180\text{-}321}$, or empty vector using an anti-HA antibody. β-actin serves as a loading control. **(C-D)** HeLa cells from (B) were infected with a titration of WT or N74D HIV-1(Luc). **(C)** Viral titre was measured 48hrs p.i. by luciferase assay. **(D)** Early reverse transcripts were measured 6hrs p.i. by qPCR. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

post-infection. We found that HA-GFP-hCPSF6[68]$_{180-321}$ was sufficient to restrict HIV-1 infection 10-fold (figure 18C), suggesting that this region of the pro-rich domain of CPSF6 is sufficient for interaction with HIV-1 CA. However, unlike HA-hCPSF6[68]-358, HA-GFP-hCPSF6[68]$_{180-321}$ did not inhibit reverse transcription (figure 18D). This shows that the RRM of HA-hCPSF6[68]-358 is dispensable for interaction with/restriction of HIV-1 but that it does influence the stage of inhibition. The RRM interacts with CPSF6's binding partner CPSF5 and also has weak RNA binding properties (discussed in section 1.5.3.1). The fact that such a small peptide of the CPSF6 pro-rich domain, which has no known function other than mediating protein-protein interactions, can inhibit the virus suggests that inhibition is mediated simply by interacting with the virus.

3.2.3 N74D mutation escapes HA-hCPSF6[68]-358 without altering infectivity

N74D mutation in CA confers resistance to inhibition by mCPSF6[72]-358 (259). Similarly, we found that N74D mutation allows escape from HA-hCPSF6[68]-358 (figures 19A and B). CA mutations often confer infectivity defects. We normalised viral input by ngRT (ng of reverse transcriptase), as measured by RT-ELISA, and found the titre of N74D and WT HIV-1 to be equivalent (figure 19C). We also tested whether N74D affected the rate of accumulation of late reverse transcripts over the first 8 hours of infection, but did not observe a significant difference (figure 19D). Given previous reports that N74D utilises a different nuclear entry pathway from WT HIV-1, we also tested whether this mutation affected the ability of the virus to enter the nucleus. However, we did not detect any difference in the ability of WT and N74D HIV-1 to form 2-LTR circles (figure 19E).

3.2.4 CPSF6 depletion does not increase viral infectivity

The hypothesis that endogenous CPSF6 restricts HIV-1 was supported by a very subtle increase in viral titre in the context of CPSF6 depletion (259). We
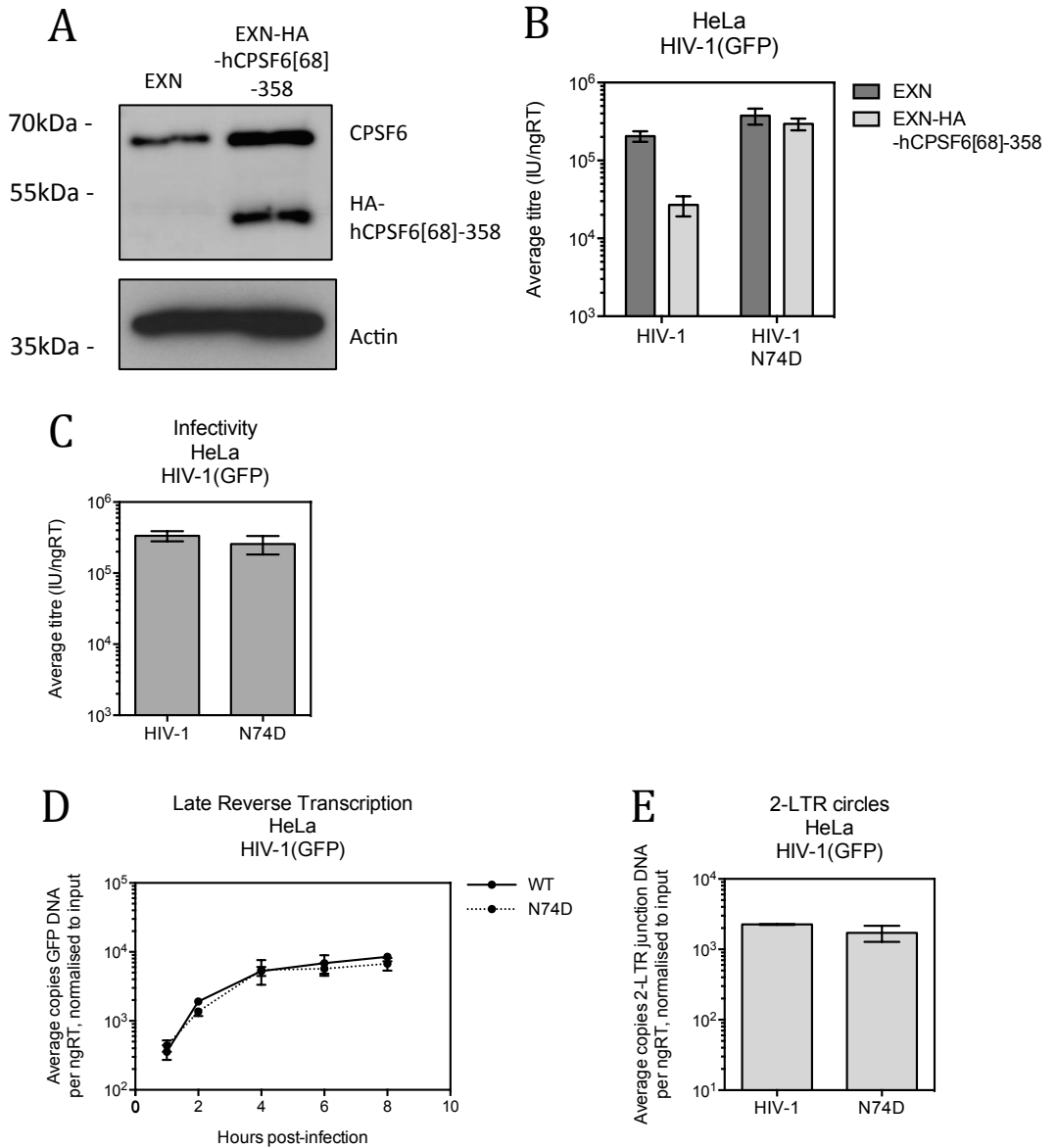
Figure 19: The N74D mutation in CA allows HIV-1 to escape restriction by hCPSF6[68]-358 without affecting viral titre

(A) A western blot of lysates from HeLa cells stably expressing HA-hCPSF6[68]-358 or empty vector using an anti-CPSF6 antibody. β-actin serves as a loading control. **(B)** HeLa cells from (A) were infected with a titration of WT or N74D HIV-1(GFP) and viral titre was measured 48hrs p.i. by flow cytometry. **(C-E)** Unmodified HeLa cells from were infected with a titration of WT or N74D HIV-1(GFP). **(C)** Viral titre was measured 48hrs p.i. by flow cytometry. **(D)** Late reverse transcripts were measured 1, 2, 4, 6, or 8hrs p.i. by qPCR. **(E)** 2-LTR circles were measured 18hrs p.i. by qPCR. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

therefore repeated this experiment and also tested the effect of depletion of CPSF5, a known binding partner of CPSF6. We transduced HeLa cells with pSIREN-based vectors expressing CPSF6 or CPSF5-targeting shRNA or a non-targeting control shRNA. CPSF6 depletion caused visible cell toxicity by 6 days post-transduction and CPSF5 by 4 days post-transduction. Therefore, these depletions were carried out transiently and with no antibiotic selection. Successful depletion of CPSF6 was detected by western blot (figure 20A) and of CPSF5 was detected by SYBR-Green qPCR (figure 20B) on the day of infection. Cells were infected with WT or N74D HIV-1(GFP) and infectivity measured 48 hours later. Whereas CPSF5 depletion had no significant affect upon viral titre, CPSF6 depletion caused a 2-3-fold decrease in viral titre (figure 20C). However, the infectivity of N74D was also reduced by a similar magnitude despite it being a CPSF6-binding mutant. Furthermore, neither CPSF5 nor CPSF6 depletion caused significant changes in accumulation of late reverse transcripts or 2-LTR circles (figures 20D and E). We therefore propose that this small defect in infectivity is due to an off-target effect of CPSF6 depletion, possibly as a result of CPSF6 depletion affecting mRNA processing (as discussed in section 1.5.3.1). Together, these data suggest that neither CPSF6 nor CPSF5 depletion significantly affects HIV-1(GFP) titre in HeLa cells. This is in agreement with our previous observation that overexpression of HA-hCPSF6[68] does not affect viral titre (figures 17B and C).

3.2.5 N74D infects independently of HIV-1 nuclear entry cofactors

As discussed in sections 1.5.1.1 and 1.5.1.2, depletion of Nup358 or TNPO3 is reported to inhibit HIV-1 infection and these proteins are therefore thought to be cofactors of HIV-1 infection. We were able to recapitulate these observations using single HeLa cell clones expressing shRNA targeting Nup358 or TNPO3, or a non-targeting control shRNA (provided by Schaller. T), which we confirmed were stably depleted by western blot (figure 21A). In our hands, the titre of HIV-1(GFP) was reduced 7-fold by Nup358 depletion and 15-fold by TNPO3 depletion (figure 21B). Neither depletion significantly affected accumulation of reverse transcripts (figure 21C) but both caused a block to 2-LTR circle
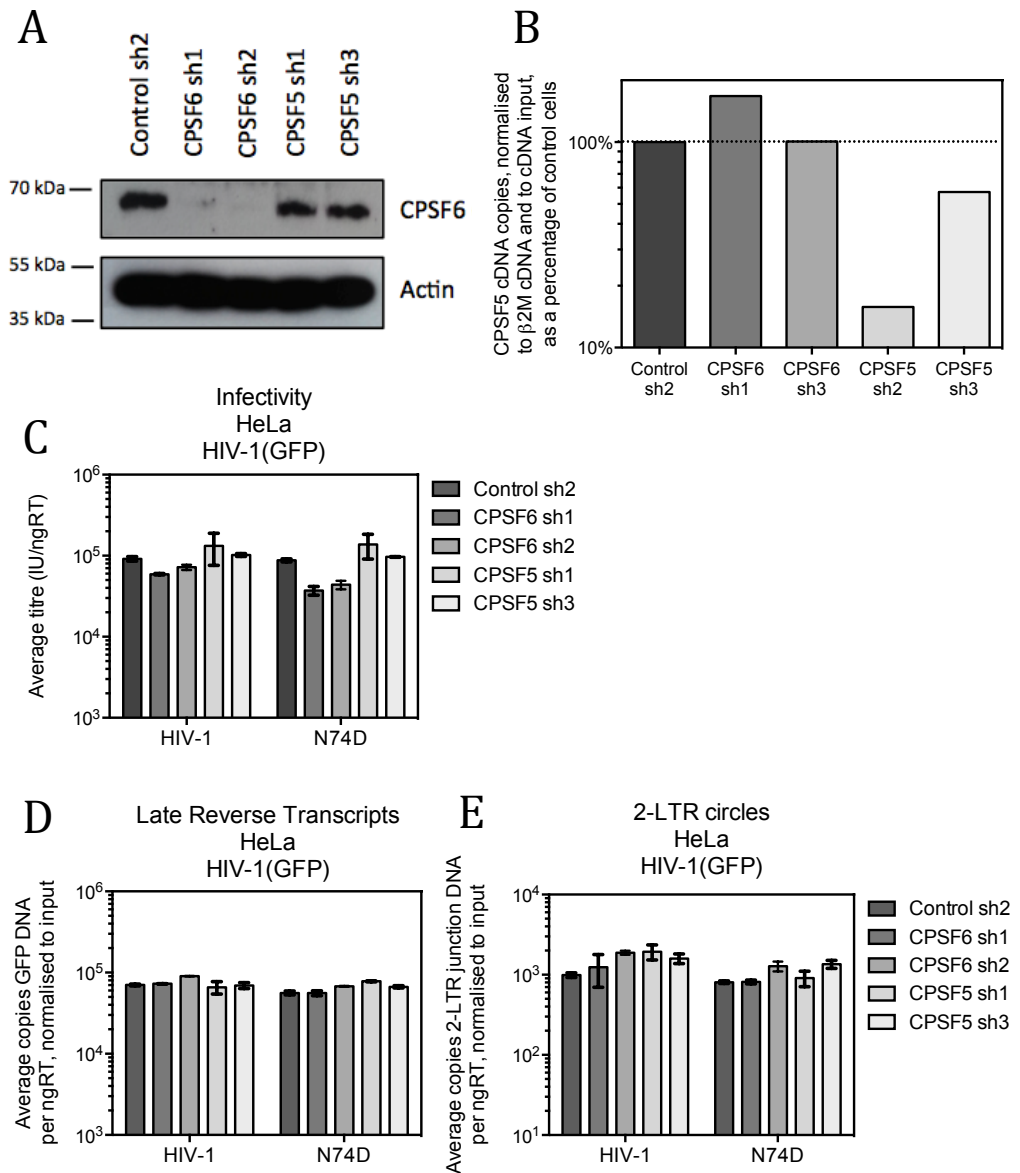
Figure 20: CPSF6 and CPSF5 depletion affect HIV-1(GFP) MFI but not infectivity, reverse transcription, or 2-LTR circle formation
(A) A western blot of lysates from HeLa cells transiently expressing CPSF6- or CPSF5-targeting shRNA or non-targeting control shRNA. β-actin serves as a loading control. **(B)** The extent of CPSF5 depletion was measured by SYBR Green RT-qPCR of CPSF5 mRNA normalised to β2-microglobulin mRNA. **(C-F)** Cells from (A-B) were infected with a titration of WT or N74D HIV-1(GFP). **(C)** Viral titre was measured 48hrs p.i. by flow cytometry. **(D)** Late reverse transcripts were measured 6hrs p.i. by qPCR. **(E)** 2-LTR circles were measured 18hrs p.i. by qPCR. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

formation that was equivalent to the block to infection (figure 21D). These findings confirm that Nup358 and TNPO3 are required for nuclear entry of HIV-1. Lee, et al. (259) reported that N74D was insensitive to depletion of these proteins. The degree of WT HIV-1 inhibition conferred by these depletions was much greater in our study and yet we still found that the infectivity of N74D was completely insensitive to these depletions (figure 21E).

In the case of Nup153, there has been dispute as to whether N74D retains sensitivity to its depletion (79, 259). We designed and tested three different Nup153-targeting shRNAs, but two were highly toxic to HeLa cells by 2 days post-transduction and so could not be used. In the case of shRNA-1, visible toxicity was not observed until day 4-5 post-transduction and so it could be used transiently to carry out infection experiments. In order to control for toxicity and to confirm that the Nup153 depleted cells were still dividing efficiently at the time of infection we infected them with the γ-retrovirus B-MLV, which cannot use nuclear pore to enter the nucleus and so requires cell division for infection (338), in parallel with WT and N74D HIV-1. Whilst B-MLV was not significantly affected by Nup153 depletion (figure 22A), infectivity of WT HIV-1 and N74D was inhibited 6-fold and 2.5-fold, respectively (figure 22B). Therefore, although N74D is significantly less dependent than WT HIV-1 upon Nup153 for nuclear entry, it retains some sensitivity to its depletion.

As well as reporting independence from these nucleoporins, Lee, et al. (259) also proposed that N74D is more dependent upon nucleoporins 85 and 155 (Nup85 and Nup155) than WT HIV-1. We transduced HeLa cells with shRNA targeting Nup85 or Nup155, or a non-targeting shRNA and selected for successfully transduced cells with puromycin before confirming their depletion by western blot (figure 22C). In our hands, Nup155 depletion did not affect the titre of either WT or N74D HIV-1. Nup85 depletion modestly but significantly reduced the viral titre of both WT and N74D (figure 22D). Importantly, there was no significant difference between the titre of WT and N74D in Nup85-depleted cells. Together, these data suggest that N74D mutation in CA causes
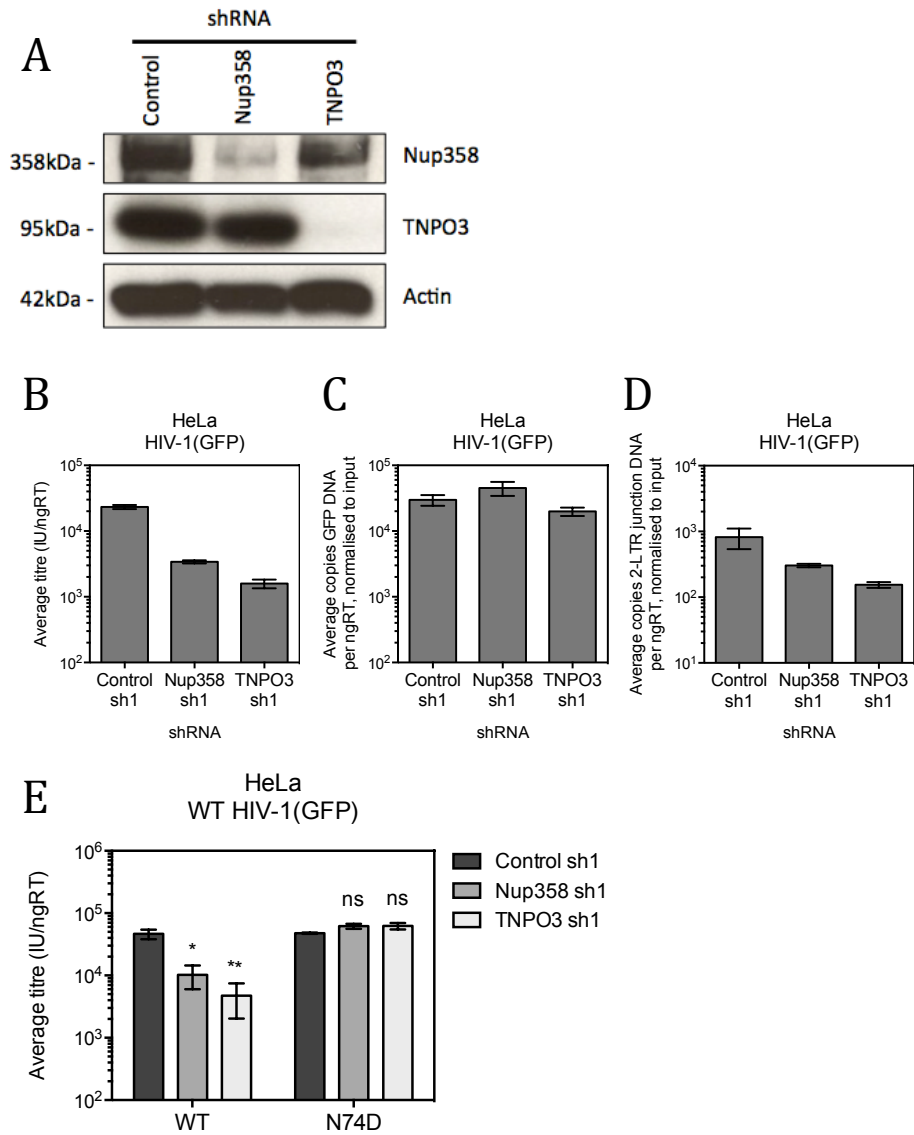
Figure 21. N74D mutation in CA makes HIV-1 insensitive to depletion of the nuclear entry cofactors Nup358 and TNPO3
(A) A western blot of lysates from HeLa cell clones stably expressing Nup358- or TNPO3-targeting shRNA or non-targeting control shRNA. β-actin serves as a loading control. **(B-D)** Cells in (A) were infected with a titration of HIV-1(GFP). **(B)** Viral titre was measured 48hrs p.i. by flow cytometry. **(C)** Late reverse transcripts were measured 6hrs p.i. by qPCR. **(D)** 2-LTR circles were measured 18hrs p.i. by qPCR. **(E)** Cells in (A) were infected with a titration of WT or N74D HIV-1(GFP) and viral titre was measured 48hrs p.i. by flow cytometry. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean). ns means p>0.05 (not statistically significant, * means p≤0.05, ** means p≤0.01, *** means p≤0.001, as measured by Student's *t*-test.
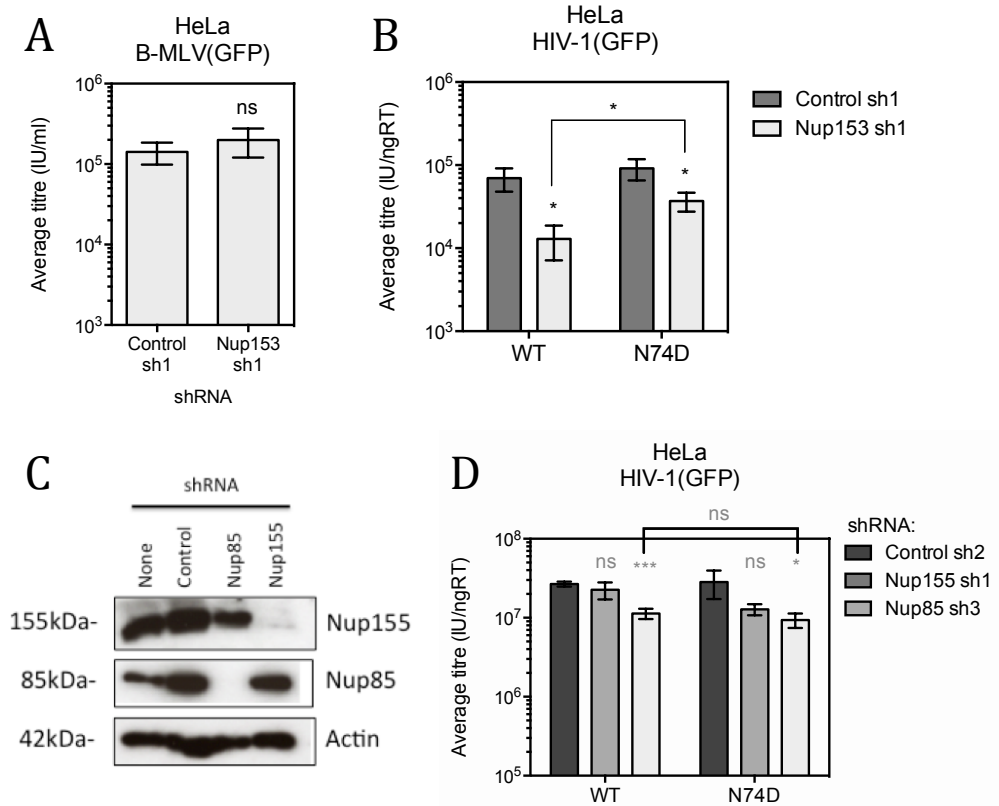
Figure 22. N74D mutation in CA reduces the sensitivity of HIV-1 to Nup153 depletion but does not affect sensitivity to Nup85 or Nup155 depletion
HeLa cells transiently expressing Nup153-targeting shRNA or non-targeting control shRNA were infected with a titration of **(A)** B-MLV(GFP) or **(B)** WT or N74D HIV-1(GFP) and viral titre was measured 48hrs p.i. by flow cytometry. **(C)** A western blot of lysates from HeLa cells transiently expressing Nup85- or Nup155-targeting shRNA or non-targeting control shRNA. β-actin serves as a loading control. **(D)** Cells in (C) were infected with a titration of WT or N74D HIV-1(GFP) and viral titre was measured 48hrs p.i. by flow cytometry. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean). ns means p>0.05 (not statistically significant, * means p≤0.05, ** means p≤0.01, *** means p≤0.001, as measured by Student's *t*-test.

HIV-1 to become less sensitive to the depletion of several HIV-1 cofactors, but that it does not gain sensitivity to any of the cofactors tested so far.

3.2.6 Depletion of CPSF6, but not of CPSF5, causes WT HIV-1 to phenocopy N74D with respect to cofactor dependence

Having found that HIV-1 infectivity is not affected by HA-hCPSF6[68] overexpression or CPSF6 depletion, there was little evidence that CPSF6 is a restriction factor or a cofactor of HIV-1 infection. However, the report that the CPSF6 binding mutant N74D is unable to replicate in MDMs (79) suggested that CPSF6 may be a cofactor in these non-permissive cells even if it is not essential for infectivity in cells as permissive as HeLa. If so, we would expect CPSF6 depletion to phenocopy N74D. In order to test whether CPSF6 depletion phenocopies N74D with respect to sensitivity to depletion of nuclear entry cofactors, we transduced HeLa cell clones stably expressing TNPO3-, Nup358-, or non-targeting shRNA (control shRNA-1), a second time such that they also expressed either CPSF6-targeting or a second non-targeting shRNA (control shRNA-2). As before, this CPSF6 depletion was done transiently to avoid toxicity. These double-depletions were confirmed by western blot (figure 23A). In cells expressing control shRNA-2, WT HIV-1 was sensitive to the depletion of either TNPO3 or Nup358 whilst N74D was unaffected by these depletions (figure 23B), in agreement with our previous findings (figure 21E). In cells depleted of CPSF6, neither WT nor N74D HIV-1 were sensitive to depletion of TNPO3 or Nup358 (figure 23C).

We similarly tested whether depletion of CPSF6 could reduce the sensitivity of WT HIV-1 to Nup153 depletion. In agreement with our previous findings (figure 22B), Nup153 depletion inhibited WT HIV-1 by 14-fold and N74D by a more modest 5-fold in cells expressing control sh2 (figure 23D). In cells depleted of CPSF6, both WT HIV-1 and N74D were inhibited approximately 4-fold by Nup153 depletion (figure 23E). Together these data show that CPSF6 determines the sensitivity of HIV-1 to depletion of Nup358, TNPO3, and Nup153. In other words, CPSF6 depletion causes WT HIV-1 to phenocopy N74D.
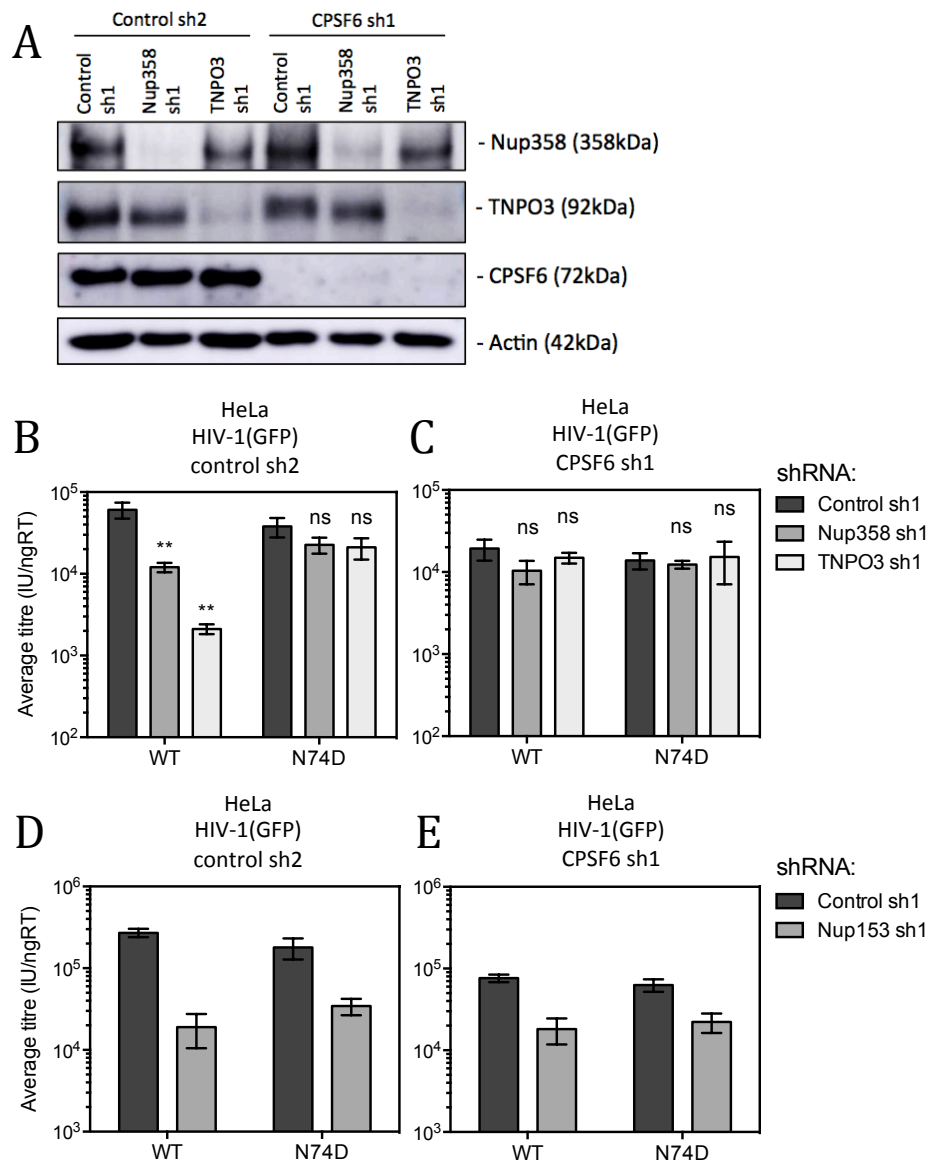
Figure 23. CPSF6 depletion alters the sensitivity of HIV-1 to depletion of nuclear entry cofactors

(A) A western blot of lysates from HeLa cell clones stably expressing Nup358- or TNPO3-targeting shRNA or non-targeting control shRNA (control sh1) and also transiently expressing CPSF6-targeting shRNA or non-targeting control shRNA (control sh2). β-actin serves as a loading control. **(B-C)** Cells in (A) were infected with a titration of WT or N74D HIV-1(GFP). Viral titre in cells transiently expressing **(B)** control sh2 or **(C)** CPSF6 sh1 was measured 48hrs p.i. by flow cytometry. **(D-E)** HeLa cells transiently expressing Nup153-targeting shRNA or non-targeting control shRNA (control sh1) and also transiently expressing CPSF6-targeting shRNA or a second non-targeting shRNA (control sh2) were infected with a titration of WT or N74D HIV-1(GFP). Viral titre in cells expressing **(D)** control sh2 or **(E)** CPSF6 sh1 was measured 48hrs p.i. by flow cytometry. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean). ns means p>0.05 (not statistically significant, * means p≤0.05, ** means p≤0.01, *** means p≤0.001, as measured by Student's *t*-test.

We interpret this to mean that there are multiple possible nuclear entry pathways that HIV-1 can take in HeLa cells and that the interaction of CA with CPSF6 dictates which pathway is taken. It seems that both pathways are equally efficient in HeLa cells, but this may not be true in less permissive cells such as MDMs.

We similarly tested whether depletion of CPSF5, CPSF6's binding partner, was able to make WT HIV-1 insensitive to depletion of TNPO3 or Nup358. Despite efficient depletion of CPSF5 mRNA, as measured by SYBR green RT-qPCR (figure 24A), HIV-1 remained sensitive to depletion of Nup358 or TNPO3 (figure 24B). In agreement with our previous finding that GFP-HA-hCPSF6[68]$_{180-321}$ is able to restrict HIV-1, this suggests that CPSF6 does not require interaction with CPSF5 in order to interact with HIV-1.

3.2.7 N74D mutation or CPSF6 depletion retargets HIV-1 integration site selection, whereas depletion of Nup153 or CPSF5 does not

Having found that CPSF6 depletion phenocopies N74D with respect to nuclear entry cofactors, we tested whether it would also phenocopy N74D with respect to integration site targeting. We also tested the effect of Nup153 depletion upon integration site selection due to the disagreement between previous reports (84, 268) and the effect of CPSF5 depletion. We infected unmodified HeLa cells with WT or N74D HIV-1. We also transiently transduced HeLa cells to express shRNA targeting CPSF6, Nup153, or CPSF5 (two CPSF5-targeting hairpins were tested due to potential problems with toxicity), or a non-targeting control shRNA and achieved efficient depletion (figures 25A and B). We infected these cells with HIV-1(GFP) at an MOI of 1.0 and extracted DNA 48hrs later. As described in section 1.2.9.2, some techniques used to amplify provirus-genome junctions can introduce bias toward certain genomic features. To avoid this, we sheared DNA randomly by ultrasonication and enriched for provirus-containing DNA using a custom designed SureSelect^{XT2} Target Enrichment Kit with baits targeting the HIV-1 LTR. This kit involves ligating barcoded adapters to each
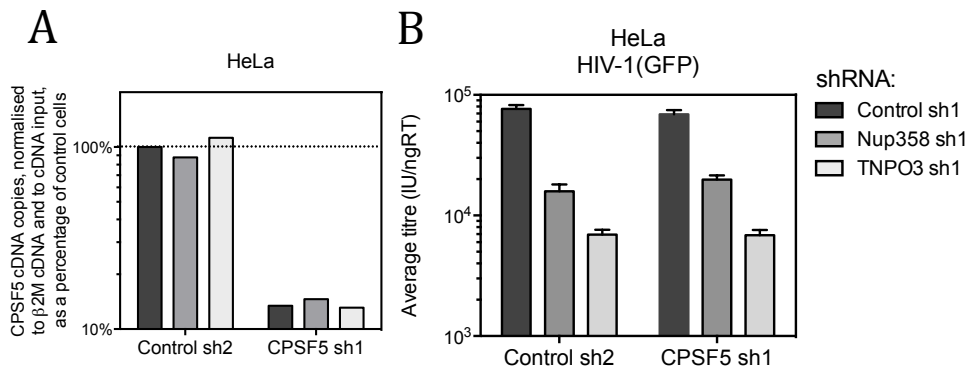
Figure 24. CPSF5 depletion does not alter the sensitivity of HIV-1 to depletion of nuclear entry cofactors

HeLa cell clones stably expressing Nup358- or TNPO3-targeting shRNA or non-targeting control shRNA (control sh1) and also transiently expressing CPSF5-targeting shRNA or non-targeting control shRNA (control sh2). **(A)** These cells were measured for the extent of CPSF5 mRNA depletion, normalised to the control β2-microglobulin, by SYBR Green RT-qPCR. **(B)** The cells in (A) were infected with a titration of WT HIV-1(GFP) and viral titre measured 48hrs p.i. by flow cytometry. All data are representative of at least 2 experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

144

A HeLa

UT   Ctrl   CPSF6

70 — CPSF6 (68kDa)
55 —
35 — Actin (42kDa)

HeLa

UT   Ctrl   N153

250 —
130 — N153 (153kDa)
55 —
35 — Actin (42kDa)

B

CPSF5 cDNA copies, normalised to β2M cDNA and to cDNA input, as a percentage of control cells

100%

10%

Untrans -duced   Control sh2   CPSF5 sh2   CPSF5 sh3

C

Name	% in gene	Genes/Mb ⬇	CpG islands/Mb
HIV-1	78.5	18.2	25.8
HIV-1 shRNA CPSF5-3	81.8	17.6	27.9
HIV-1 shRNA Nup153	75.3	17.6	26.0
HIV-1 shRNA control	78.2	17.6	23.4
HIV-1 shRNA CPSF5-2	71.7	17.3	27.2
N74D	70.4	7.0	6.1
HIV-1 shRNA CPSF6	62.5	6.4	5.6

D

au bp
edge #

Height

0.03
0.02
0.01
0.00

100 99
1

HIV-1 N74D   CPSF6 sh1

100 99
5

95 48
4

98 44
3

95 59
2

CPSF5 sh2   Control sh2   CPSF5 sh3   HIV-1 WT   N153 sh1

Figure 25. N74D mutation and CPSF6 depletion, but not CPSF5 or Nup153 depletion, alter HIV-1 integration site selection

HeLa cells transiently expressing CPSF6-, Nup153-, or CPSF5-targeting shRNA or a non-targeting control shRNA were measured **(A)** for the extent of CPSF6 or Nup153 depletion by western blot and **(B)** for the extent of CPSF5 mRNA depletion, normalised to the control β2-microglobulin, by SYBR Green RT-qPCR. **(C-D)** Unmodified HeLa cells were infected with WT or N74D HIV-1(GFP) at an MOI of ~1.0. Cells in (A) and (B) were infected with WT HIV-1(GFP) at an MOI of ~1.0. At 24hrs p.i. DNA was extracted, sheared, enriched for fragments containing HIV-1 LTRs using the SureSelect[XT2], and sequenced. **(C)** The results of integration site analysis for gene density, CpG-island density, and the proportion of proviruses within genes, sorted by gene density as indicated by the black arrow. **(D)** A dendrogram based upon the gene density and CpG island density values shown in (C). The y-axis of the dendrogram, Height, uses arbitrary units to represent the relative closeness of either individual data points or of clusters. The nodes are statistically supported by bootstrap probability (bp) *p* values shown in green and approximately unbiased (AU) *p* values shown in red. The red/blue shading in C and D highlights the two major clusters formed by these samples.

DNA sample. This allowed us to combine multiple samples before sequencing DNA by MiSeq (Illumina) using primers designed against these adapters.

Computational analysis was subsequently used to characterise the DNA surrounding integration sites (all computational analysis was carried out by Riccombeni A). Reads were first mapped to the pCSGW-derived provirus using Bowtie2, and any sequences containing solely proviral DNA were discarded. Where sequences partially mapped to the provirus, the non-viral DNA was mapped to the human genome (version hs19) using tophat2. Identical integration sites were assumed to be PCR duplicates and so only one copy was kept in the library. Reads mapping to multiple regions of the genome were also discarded. Between 206 and 2606 individual integration sites were identified for each sample. After obtaining genomic coordinates of each integration site, the proportion of sites lying within a gene were measured. Furthermore, gene density and CpG island density (according to RefSeq annotation of hs19) were measured in a 1Mbp window surrounding each integration site (500Kbp either side of the provirus) (values shown in figure 25C). Gene and CpG island density were used to cluster the samples using Pvclust and the resulting dendrogram is shown in figure 25D. The bootstrap probability (BP) and approximately unbiased (AU) (AU assesses the variation in BP values when changing sample size, which helps to reduce the bias in BP values caused by sample size (339)) *p*-values shown on the dendrogram were then calculated to assess the confidence in tree selection.

Similarly to Schaller *et al*. (79), we found that WT HIV-1 and N74D integrated into regions with average gene densities of 18.2 and 7.0 genes/Mbp, respectively. As hypothesised, CPSF6 depletion caused HIV-1 to integrate into regions of a similar gene density to N74D in unmodified cells, 6.4 genes/Mbp. In contrast, expression of a control shRNA caused HIV-1 to integrate into regions of a similar gene density to that of HIV-1 in unmodified cells, 17.6 genes/Mbp. Measurements of CpG island density followed the same pattern; in unmodified cells and cells expressing non-targeting shRNA HIV-1 integrated into regions of 25.8 and 23.4 CpG-islands/Mbp, respectively, whilst in cells depleted of CPSF6

HIV-1 integrated into regions of 5.6 CpG-islands/Mbp, thereby phenocopying N74D, which integrated into unmodified cells in regions of 6.1 CpG-islands/Mbp. Based on these data, WT HIV-1 and control sh2 cluster together, whilst N74D and CPSF6 sh1 cluster together. In contrast to CPSF6, CPSF5 sh2 and 3 both cluster with control sh2, indicating that they did not have a significant effect upon gene density or CpG island density at integration sites.

We were interested to test whether Nup153 depletion affects integration site targeting because although it is a cofactor of HIV-1 infection, its use is only partially dictated by CPSF6. In cells efficiently depleted of Nup153, HIV-1 integrated into regions of 17.6 genes/Mbp and 26.0 CpG-islands/Mbp, causing it to cluster with control cells. This suggests that although Nup153 is required for nuclear entry of HIV-1, it does not influence HIV-1 integration site targeting in the same way as TNPO3 and Nup358.

Preferentially integrating within genes is a feature of HIV-1 attributed to the interaction between IN and LEDGF/p75 (153, 154) (discussed in section 1.2.9.3). However, it has been hypothesised that mutations in CA determine a downstream interactions between IN and HIV-1 cofactors (250) and so it was important to test whether these depletions/CA mutations affected the proportion of proviruses lying within genes. However, in all the conditions tested the virus still preferentially integrated inside genes, which suggests that neither CPSF6, CPSF5, nor Nup153 affect downstream interaction with LEDGF/p75.

3.3 Discussion

Since these data were obtained, a great deal more research into the role of CPSF6 in HIV-1 infection has been published that influences our interpretation of these data, which will be discussed here. Crucially, the interaction between CPSF6 and CA has been shown to be direct. The crystal structure of the 15-residue peptide CPSF6[68]$_{276-290}$ (amino acid sequence PVLFPGQPFGQPPLG) in complex with both monomeric CA$_{NTD}$ and hexameric CA has now been solved (340-342). These structures revealed that CPSF6 interacts with CA in a pocket that extends across two adjacent monomers within a hexamer, as shown in figure 26A. Therefore, CPSF6[68]$_{276-290}$ interacts with hexameric CA with a 14-fold higher affinity than it does with monomeric CA. Figure 26B shows that the CPSF6 binding site includes NTD helices 3 and 4 of one CA monomer (called the 'first' monomer) but also NTD helices 2 and 7 and CTD helices 8 and 9 of the adjacent monomer (called the 'second' monomer). Figure 26C highlights the residues of both proteins that form key interactions, which includes residue N74, as expected. CA residues N74 and N57 from the first monomer form hydrogen bonds with CPSF6 residues L278 and F284, respectively, whilst CA residues K182 and Q179 from the second monomer form hydrogen bonds with CPSF6 residues G281 and P280, respectively (340). Like N74D (259), several other mutations in this CA pocket including N57A, M66F, Q67A, K70A, S102D, and K182R, have been shown to confer escape from C-terminal truncations of CPSF6 (83, 340, 342). The peptide CPSF6[68]$_{276-290}$ is hydrophobic and buries into this CA pocket, forming an almost closed loop. Its termini both project out of the binding pocket, and so we speculate that this region of CPSF6 may protrude from the full-length protein in order to interact with CA. However, no crystal structure of full-length CPSF6, or even of the pro-rich domain, has been solved as yet.

hCPSF6[72]-358 restriction assays and ITC have been used to show that the interaction between CA and CPSF6 is conserved within primate Lentiviruses (HIV-1, HIV-2, SIVmac, SIVmne) but not in other Lentivirus serogroups (FIV, EIAV, BIV) or γ-retroviruses (MLV) (83, 259, 342). Figure 27A shows is an alignment of CA amino-acid sequences from 7 different primate Lentiviruses,
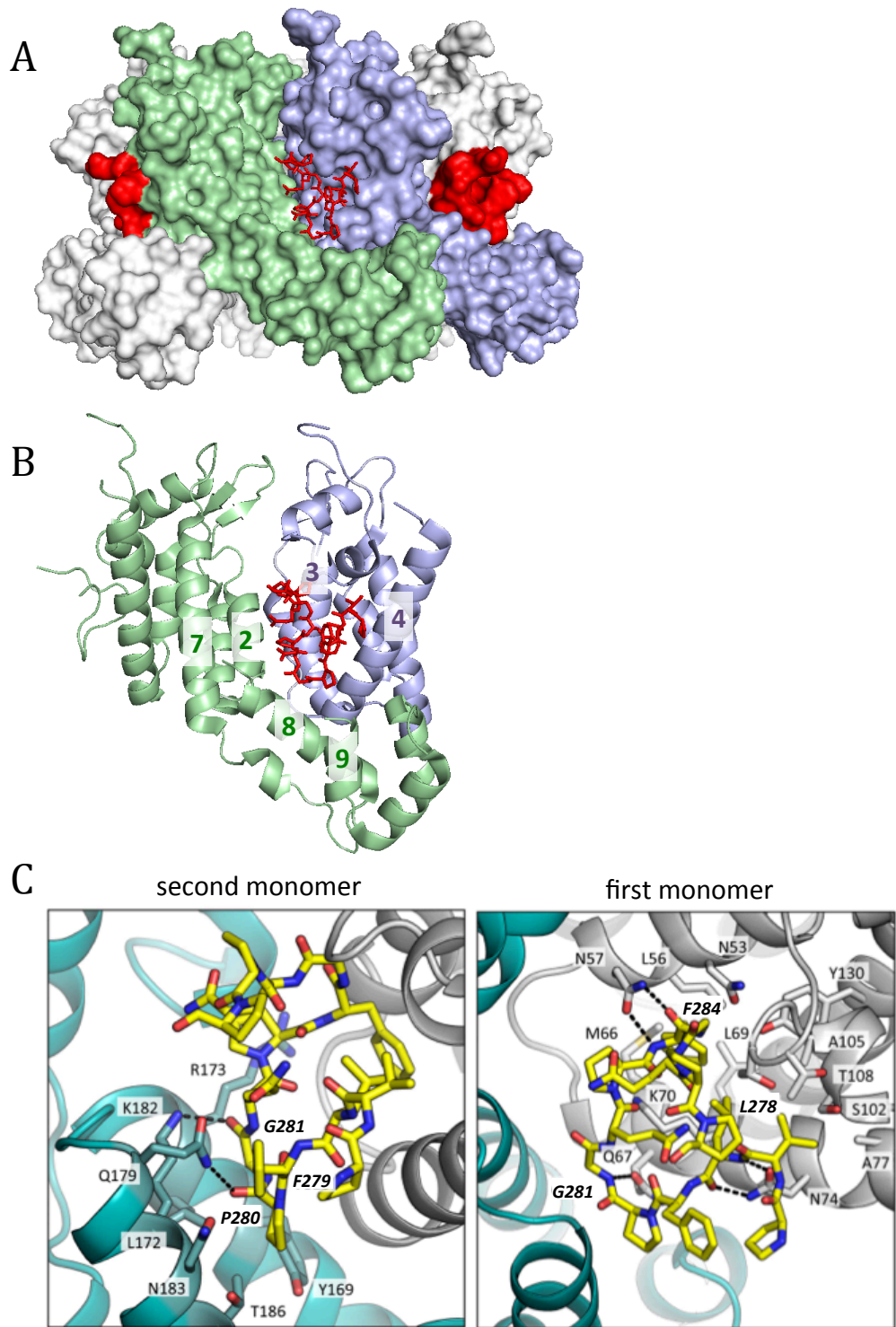
Figure 26. CPSF6 interacts with two adjacent monomers in a CA hexamer
(A) Crystal structure of CPSF6[68]$_{276-290}$ in complex with hexameric CA (PDB 4U0A from Price et al. 2014). Peptides in red, first CA monomer in purple, second CA monomer in green, all other CA monomers in grey. **(B)** As in (A) but only showing the second CA monomer and the CA$_{NTD}$ of the first monomer. Numbers highlight helices of CA bound by the peptide. **(C)** Crystal structure of CPSF6[68]$_{276-290}$ in complex with hexameric CA. Peptide in yellow/red/blue, first monomer in grey, second monomer in teal. Interacting residues are labeled, where bold labels indicate residues forming hydrogen bonds. Image adapted from Price et al. 2014.

Figure 27. The CPSF6 binding pocket of CA is conserved in primate Lentiviruses

(A) An alignment of CA amino acid sequences from seven primate Lentiviruses including HIV-1 M group and O group, HIV-2, and four SIV strains. Red or blue arrows indicate CA residues from first or second monomer, respectively, that interact with CPSF6[68]$_{276-290}$. Residues that form hydrogen bonds with CPSF6[68]$_{276-290}$ are labeled with text. (*) indicates completely conserved residues, (:) indicates conservative mutations, (.) indicates semi-conservative mutations, and no symbol indicates no conservation at this position. **(B)** Crystal structure of CPSF6[68]$_{276-290}$ in complex with hexameric CA (PDB 4U0A from Price et al. 2014) showing conservation of residues in primate Lentiviruses, from completely conserved in red through to unconserved in pale yellow. CPSF6 peptide in cyan.

which reveals that 20 of the 24 CA residues mediating interactions with CPSF6[68]$_{276-290}$ are completely conserved within these primate Lentiviruses. When the level of conservation is mapped onto a CA hexamer, as in figure 27B, it can be seen that the CPSF6 binding pocket is one of the most conserved regions of the protein. Similarly, it has been shown that the CA-binding region of CPSF6 is completely conserved in primate orthologs spanning 35million years of evolution (343). Whilst these observations in themselves are not evidence for or against any hypothesised role of CPSF6, they are consistent with CPSF6 being a cofactor of HIV-1 infection.

3.3.1 The interaction of CA and endogenous CPSF6 does not affect HIV-1 infectivity in HeLa cells

Is discussed in section 1.5.3.3, Lee *et al.* support their hypothesis that CPSF6 is a restriction factor of HIV-1 with the findings that hCPSF6[72] overexpression inhibited HIV-1 infection by 2-fold and that depletion of endogenous CPSF6 increased HIV-1 infectivity by a modest but statistically significant 1.2-fold (259). In disagreement with this, we found that neither overexpression of HA-hCPSF6[68] nor depletion of endogenous CPSF6 significantly altered HIV-1 infectivity. Other studies have since recapitulated our observations (247, 344, 345). Furthermore, we found that N74D mutation did not affect the infectivity of HIV-1, the rate of reverse transcription, or the ability of the virus to enter the nucleus and this has also been reproduced by other groups (344, 345). Whilst these observations argue against CPSF6 being a restriction factor, they also argue against it being a cofactor of HIV-1 infection in HeLa cells.

3.3.2 Cytoplasmic forms of CPSF6 restrict HIV-1 by simply binding to CA

In light of the finding that CA and CPSF6 interact directly, there has been great interest in the mechanism of inhibition of mCPSF6[72]-358, despite it being an artificial protein. Lee *et al.* (259) noted that both of the antiviral factors they created (mCPSF6[72]-358 and mCPSF6[72]-526) had lost their classical NLS

and were more abundant in the cytoplasm than endogenous CPSF6, which is predominantly nuclear. It was hypothesised that C-terminal truncations of CPSF6 are antiviral because they mislocalise to the cytoplasm, rather than because they have lost some other function. To test this hypothesis, Fricke *et al.* (345) fused the prototypic classical NES from protein kinase inhibitor α (PKIα) to the N-terminus of full-length hCPSF6[68], creating NES-hCPSF6[68]. As hypothesised, this protein was almost entirely cytoplasmic and potently inhibited HIV-1. Correspondingly, it was shown that fusion of the prototypic classical SV40 NLS to the C-terminus of hCPSF6[68]-358 (342) or hCPSF6[72]-358 (247) resulted in them localising exclusively to the nucleus and losing all antiviral activity.

Interestingly, two studies have assessed localisation of HA-hCPSF6[68]-358 through immunofluorescence using an anti-HA antibody (as opposed to an anti-CPSF6 antibody which confusingly detects endogenous CPSF6 as well). Whilst they did observe a significant proportion of this protein in the cytoplasm, the majority remained in the nucleus (247, 342), despite both predicted NLSs having been removed. This might be because CPSF6 has another unknown NLS or because it enters the nucleus in complex with another protein. Either way, the amount of truncated CPSF6 in the cytoplasm is evidently sufficient to inhibit HIV-1.

Having established that CPSF6 must be able to interact with CA and must be cytoplasmic in order to inhibit HIV-1, the mechanism of restriction was questioned. In this study, we have shown that a GFP-HA-hCPSF6[68]$_{180-321}$ is able to inhibit HIV-1. Residues 180-321 are 44% proline and have no known function, but are predicted to mediate protein-protein interactions and contain the CA-binding region. Since making this observation, it has been reported that an even smaller region of CPSF6 fused to GFP – GFP-CPSF6[72]$_{301-358}$ – is also able to inhibit HIV-1 (343). The fact that such small regions of CPSF6 can inhibit the virus suggests that no functional domain of CPSF6 is required and that simply binding to CA may be sufficient to mediate restriction.

Whilst the field concurs that cytoplasmic localisation and an ability to bind to CA are two features required by CPSF6 C-terminal truncations to restrict HIV-1, the mechanism by which this restriction occurs remains debated. Given that depletion of CPSF6 has no effect on HIV-1 titre, CPSF6 C-terminal truncations cannot simply be displacing endogenous CPSF6; they must have some direct effect upon viral cores. Several studies have now reported that CPSF6 (and derivatives of CPSF6) affects the stability of HIV-1 cores. Consistent with this hypothesis, CPSF6[68]$_{276-290}$ interacts with residue L172, which is involved in interhexamer CTD-CTD interactions, and residues K70 and K182, which are involved in intrahexamer NTD-CTD interactions (346). Whilst most studies have found that CPSF6 stabilises cores (90, 247, 345) it has also been reported that CPSF6 destabilises cores (344). Core stability has been studied through two different biochemical assays, the CA-NC tube assay and the Fate of Capsid (FoC) assay. The CA-NC tube assay is a purely *in vitro* assay. Recombinant CA-NC fusion proteins are produced in *E.coli* and, under the correct pH and salt concentrations and in the presence of RNA, they spontaneously assemble into hollow helical tubes of varying lengths (347, 348). CA-NC tubes resemble HIV-1 cores because they are formed from a hexameric CA lattice. These tubes are incubated with the recombinant protein being tested and are then centrifuged through a sucrose cushion. Whereas intact CA-NC tubes will be pelleted, disassembled CA-NC will not. The proportion of pelletable CA can be measured by western blot and used to assess the degree of disassembly. This technique has been used to show that recombinant hCPSF6[68]-321 stabilises CA-NC tubes (90). This assay has been widely criticised for two main reasons. Firstly, CA-NC tubes differ from viral cores in shape, size, and composition, meaning it is unclear whether they have the same stability properties as HIV-1 cores. Secondly, the CA-CPSF6 interaction does not occur in the context of a cell. To avoid these two issues, the FoC assay was developed (114). This assays involves infecting cells that express the protein of interest and then centrifuging cell lysates through a sucrose cushion. The proportion of pelletable intact CA is then measured by western blot in the same way as the CA-NC tube assay. The FoC assay has been used to show that mCPSF6[72]-358 (247) and NES-CPSF6[68] (345) stabilise cores, which is in agreement with the data obtained from the CA-

NC assay. In disagreement with these findings, one group used the FoC assay to show that CPSF6 C-terminal truncations including exon 6 have no effect upon core stability but those excluding exon 6 induce premature uncoating (344). However, this observation has not yet been repeated.

A caveat of all biochemical assays of core stability is that they measure changes in the stability of all the viral cores present in an experiment and cannot differentiate between cores that would have gone on to be successfully infectious and those that would not. This is an important flaw because it is established that the vast majority of virions do not successfully reach integration and are thus not infectious. Upper estimates of the ratio of infectious to non-infectious HIV-1 virions reach 1:1000, although these figures differ depending on the strain of the virus and the technique used to produce the virus (349, 350). This disparity is thought to be partly due to the production of inherently defective virions but also partly because some virions simply, by chance, do not come into contact with required host cofactors. Virions can be defective at many different stages of infection. However, the fact that cell-to-cell transmission is so much more efficient than cell-free transmission suggests that entry in particular is a high barrier to infection (61). In agreement with this, Thomas, et al. (102) showed that the majority of non-infectious virions fail to reach the stage of reverse transcription and calculated that, of the virions that do successfully reverse transcribe, 1 in 8 is infectious. It has therefore been argued that the FoC assay, which only takes into accounts cores that have entered a cell, is more representative of infectious virions than purely *in vitro* assays. However, even this lower ratio of 1:8 still means that the majority of virions in the FoC assay do not represent infectious virions.

Despite the aforementioned issues with biochemical stability assays, there is a consensus that cytoplasmic/antiviral forms of CPSF6 stabilise HIV-1 cores. We therefore asked why endogenous CPSF6 does not also inhibit cores. One possibility is that a certain CA occupancy must be achieved by CPSF6 (i.e. a certain proportion of the ~1056 CA monomers within a core must be bound by CPSF6) in order to affect stability and that cytoplasmic concentrations of

endogenous CPSF6 are too low to reach this threshold. An alternative possibility is that the low levels of cytoplasmic endogenous CPSF6 are sufficient to stabilise cores but that this stabilisation is transient because endogenous CPSF6 is only transiently in the cytoplasm before rapidly shuttling back into the nucleus (as discussed in section 1.5.3.1).

3.3.3 Antiviral forms of CPSF6 can inhibit HIV-1 at different stages of infection

In contrast to mCPSF6[72]-358, which reportedly inhibits after reverse transcription but before HIV-1 nuclear entry (259), we found that HA-hCPSF6[68]-358 inhibits reverse transcription. This implicated exon 6 in the ability of C-terminal CPSF6 truncations to inhibit reverse transcription. In agreement with this, Hori *et al.* (344) recently showed that removal of exon 6 from several different C-terminal CPSF6 truncations (both human and murine) resulted in the ability to inhibit reverse transcription. However, the RS domain has also been implicated in the stage of inhibition by the finding that NES-hCPSF6[68] inhibits after reverse transcription (345). Furthermore, the RRM has been implicated in the stage of inhibition by our finding that HA-GFP-hCPSF6[68]$_{180-321}$ inhibits after reverse transcription.

Whilst it is possible that exon 6, the RRM, and the RS domain each have a specific and independent effect upon the virus, we consider this unlikely. As discussed in section 1.2.7.2, there is evidence, albeit controversial, that stabilising cores can inhibit reverse transcription. Therefore, one possibility is that some CPSF6 truncations stabilise cores more than others and that these are the forms of CPSF6 that inhibit reverse transcription. Models of the intact core structure have suggested that the core is porous enough to allow dNTPs to access the viral genome, allowing reverse transcription to proceed inside the core. Forms of CPSF6 that inhibit reverse transcription may physically obstruct these pores and/or alter their conformation such that dNTPs cannot enter the core. Either way, it is important to note that NES-CPSF6[68], the construct most

likely to represent endogenous CPSF6, does not inhibit reverse transcription (345).

3.3.4 The CA-CPSF6 interaction determines the nuclear entry pathway of HIV-1

We, and others, have now shown that the N74D mutation in CA causes HIV-1 to lose sensitivity to depletion of Nup358 and TNPO3 (figure 21E) (79, 259). N74D mutation has been reported to reduce the affinity of CA for Nup358$_{Cyp}$ by 6-fold (79) and so it could be suggested that N74D is insensitive to depletion of Nup358 simply because it cannot interact with it. However, if this were true, we would expect N74D mutation to cause a similar infectivity defect as Nup358 depletion does, which we have shown is not the case. We therefore hypothesised that the insensitivity of N74D to Nup358 and TNPO3 depletion was due to its inability to interact with CPSF6, as opposed to another unknown phenotype of this mutant. In support of this hypothesis, several other CPSF6-binding CA mutants have since been shown to infect independently of Nup358 and TNPO3, including N57A, Q67A, K70R, and T107A (342). In this study, we confirmed this hypothesis by showing that CPSF6 depletion causes WT HIV-1 to phenocopy N74D and become insensitive to depletion of Nup358 or TNPO3. In other words, the dependence of HIV-1 upon Nup358 and TNPO3 is dictated by the upstream interaction of CA with CPSF6.

Two studies have since reproduced the finding that HIV-1 is insensitive to TNPO3 depletion in the context of CPSF6 depletion (247, 345). However, their interpretation of this finding differs from ours. De Iaco, et al. (247) postulated that TNPO3 (a karyopherin or SR proteins) imports CPSF6 (an SR protein) into the nucleus. They used immunofluorescence to show that TNPO3 depletion results in a slight increase in the cytoplasmic levels of endogenous CPSF6. In support of this idea, Maertens et al. (351) reported that TNPO3 interacts directly with CPSF6 and genetically mapped this interaction to the RS domain of CPSF6. In contrast, Fricke, et al. (345) did not detect a change in CPSF6 localisation after TNPO3 depletion, despite their TNPO3 depletion being

sufficient to relocalise positive control proteins. Based on their observations, De Iaco *et al.* suggested that the only reason TNPO3 depletion inhibits HIV-1 is because it causes CPSF6 to accumulate in the cytoplasm, which is antiviral. Importantly, even if TNPO3 does transport CPSF6 into the nucleus, this does not negate that possibility that TNPO3 also imports IN into the nucleus or that CA-CPSF6 interaction determines the downstream TNPO3-IN interaction, as we have hypothesised.

Whilst N74D is insensitive to depletion of Nup358 or TNPO3, we found this CA mutant to be sensitive to Nup153 depletion, albeit significantly less sensitive than WT HIV-1. Intriguingly, it has recently been reported that $Nup153_{1410-1417}$ interacts directly with CA in the same region of CA as $CPSF6[68]_{276-290}$, a pocket that spans two adjacent monomers within a CA hexamer (as shown in figure 28A) (340). CPSF6 F284 and Nup153 F1417 lie in almost identical positions and form hydrogen bonds with CA N57 that is essential for interaction (figure 28B). Despite many similarities, the CPSF6 and Nup153 peptides do make distinct interactions with CA and, as shown in figure 28C, Nup153 does not hydrogen bond with CA residue N74 (340). Accordingly, it has been shown that N74D mutation does not affect the ability of $Nup153_{CTD}$ to co-immunoprecipitate with CA_{NTD} (83). Therefore, the reduced sensitivity of N74D to depletion of Nup153 cannot be due to N74D CA being unable to interact with it and hypothesised that, like Nup358 and TNPO3, dependence upon Nup153 is determined by CA-CPSF6 interaction. Consistent with this, we found that CPSF6 depletion caused WT HIV-1 to phenocopy N74D with respect to Nup153 depletion (as with Nup358 and TNPO3 depletion).

We were intrigued by the observation that HIV-1 retains moderate sensitivity to Nup153 depletion in the context of CPSF6 depletion or N74D mutation. The CA-Nup153 interaction is conserved within primate Lentiviruses and Nup153 depletion inhibits all primate Lentiviruses tested thus far by at least 10-fold (83, 340). We interpret these observations to mean that Nup153 depletion inhibits WT HIV-1 directly because Nup153 is a specific cofactor of infection. Interestingly, Nup153 depletion also inhibits the non-primate Lentiviruses FIV
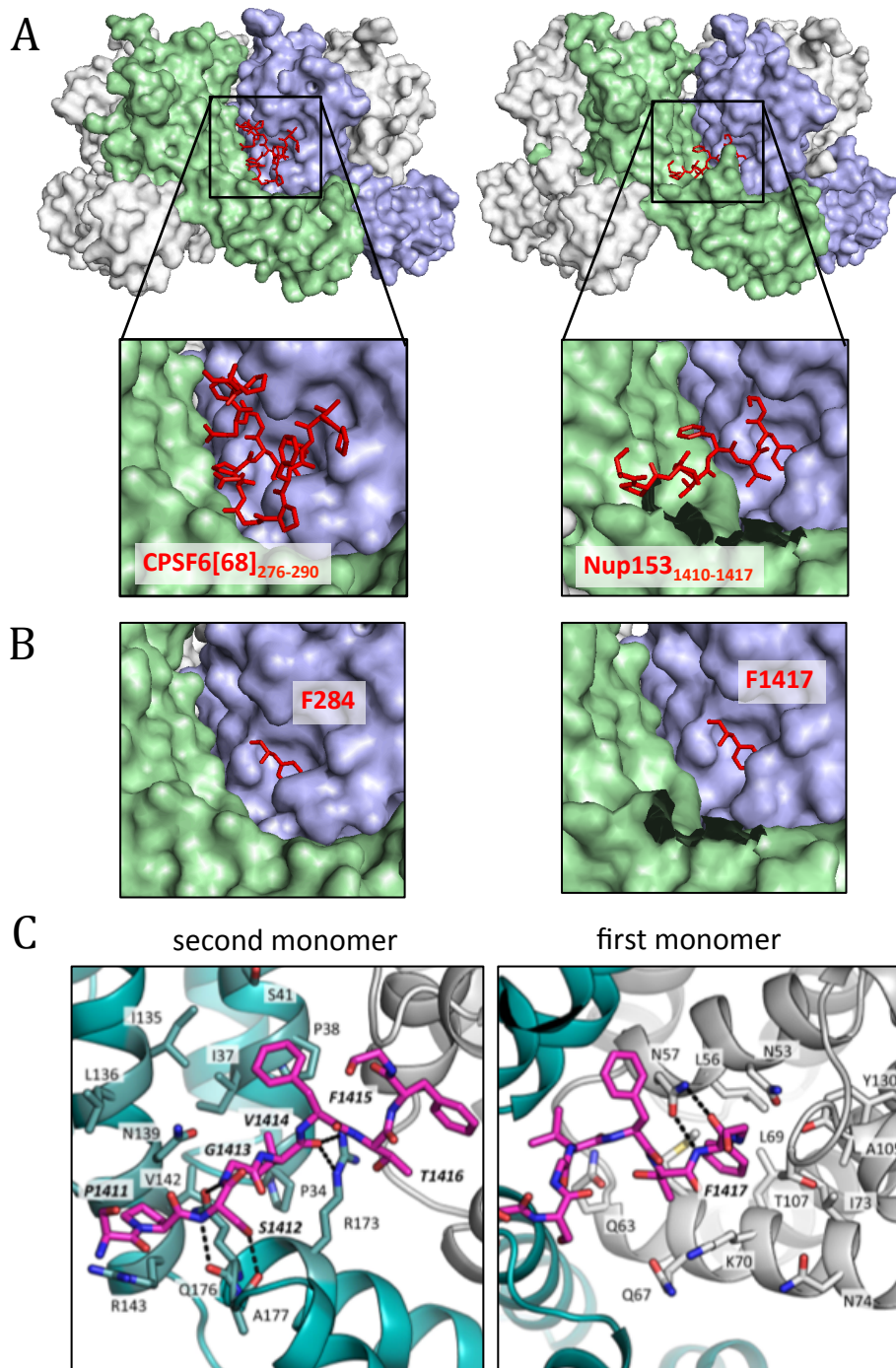
Figure 28. Nup153 interacts with CA in the same pocket as CPSF6
(A) Crystal structures of CPSF6[68]$_{276-290}$ (left) or Nup153$_{1410-1417}$ (right) in complex with hexameric CA (PDBs 4U0A and 4U0C, respectively, both from Price et al. 2014). Peptides are in red, first CA monomers in purple, second CA monomers in green, and all other CA monomers in grey. **(B)** A magnified view of (A) showing only CPSF6 F284 and Nup153 F1417, which interact with CA N57 in almost the exact same manner. **(B)** Crystal structure of Nup153$_{1410-1417}$ in complex with hexameric CA. Peptide in pink/red/blue, first monomer in grey, second monomer in teal. Interacting residues are labeled, where bold labels indicate residues forming hydrogen bonds. Image adapted from Price et al. 2014.

and BIV by ~3-fold despite the fact that it cannot interact with their CAs (83). As discussed in section 1.5.1.3, Nup153 forms the nuclear basket and is required to anchor the NPC into the plasma membrane. Its depletion therefore has dramatic effects on the overall structure of the NPC (261). We therefore propose that Nup153 depletion inhibits FIV and BIV indirectly, by generally disrupting the structure of the NPC (which these viruses require to enter the nucleus (9)). Although the route taken by N74D to the nucleus is unclear, it has been shown to efficiently infect non-dividing cells and so must be able to access the nucleus via the NPC (79). We therefore hypothesise that N74D remains moderately dependent upon Nup153 because, like FIV and BIV, it requires an intact NPC to enter the nucleus. In other words, N74D mutation relieves HIV-1 of dependence upon CA-Nup153 interaction but it does not relieve the virus of dependence upon the NPC.

Taken together, these observations suggest that CA-CPSF6 interaction causes the virus to infect via a nuclear entry pathway that is dependent upon specific interaction with Nup358, TNPO3, and Nup153. Without CA-CPSF6 interaction, the virus takes an alternative – but equally efficient – nuclear entry pathway that is independent from specific interaction with these cofactors but remains dependent upon the NPC. Lee *et al.* suggested that N74D mutation caused HIV-1 to become sensitive to depletion of Nup155 and Nup85. However, we were unable to reproduce this observation. Instead we found that neither WT nor N74D HIV-1 was sensitive to Nup155 depletion and that both were equally sensitive to Nup85 depletion. Therefore, the alternative route taken by N74D remains uncharacterised. The modest sensitivity of N74D to depletion of Nup153 and Nup85 suggests that it requires an intact NPC to enter the nucleus, but no specific interactions with cofactors have been identified as yet. It is possible that N74D reverse transcripts enter the nucleus via a relatively non-specific pathway, such as transfection.

It was recently reported that, like CypA-binding mutations (discussed in section 1.5.2.5), the CPSF6-binding CA mutations N74D and N57S confer resistance to MxB, a recently identified restriction factor of HIV-1 nuclear entry (307, 308). It

therefore appears that CA-CPSF6 interaction not only dictates the dependence of HIV-1 upon downstream nuclear entry cofactors but also the sensitivity of the virus to downstream restriction factors.

There is substantial evidence that both CA-CypA interaction and CA-CPSF6 interaction determine downstream interactions with Nup358, TNPO3, Nup153, and MxB. We hypothesise that both CypA and CPSF6 stabilise cores and prevent their uncoating and that CPSF6, a protein that is present in the cytoplasm but rapidly shuttles to the nucleus (322), traffics cores to NPCs. As discussed in section 1.2.7.1, intact cores have been visualised docked at NPCs. Of course, if cores are to uncoat at the NPC and enter the nucleus, CypA and CPSF6 must dissociate from them. Given that Nup358 interacts with the same CA loop as CypA and Nup153 interacts with the same CA pocket as CPSF6, we speculate that Nup358 and Nup153 compete CypA and CPSF6, respectively, off cores once they reach the NPC. If neither Nup358 nor Nup153 stabilises cores – something that has not yet been determined – cores could then uncoat in the vicinity of the NPC. After uncoating, the viral IN would be exposed to TNPO3, which transports the viral PIC into the nucleus. If CA cannot interact with CypA or CPSF6 (either due to CA mutation or protein depletion), we hypothesise that cores uncoat prematurely in the cytoplasm. The viral PIC (minimally including viral DNA and IN) is able to enter the nucleus through an alternative pathway, which remains uncharacterised but is independent of specific interaction with Nup358, TNPO3, and Nup153.

3.3.5 The CA-CPSF6 interaction determines the integration site targeting of HIV-1

As discussed in section 1.2.9.3, HIV-1 preferentially integrates inside genes due to the interaction between IN and LEDGF/p75 (153). HIV-1 also integrates into regions that are enriched in genes and transcriptionally active (138, 140). It has been shown that depletion of Nup358 or TNPO3 causes HIV-1 to integrate into regions of a gene density that is significantly lower than in control cells, but is still enriched in genes compared to a random site in the genome (159).

Furthermore, N74D HIV-1, which infects independently of both TNPO3 and Nup358, integrates into regions of an even lower gene density that is equivalent to a random site in the genome (79). In other words, this CPSF6-binding CA mutant has completely lost its ability to selectively integrate into gene-rich regions. In this study we used an unbiased methodology to sequence and characterise HIV-1 integration sites. Using this technique we recapitulated the observation that N74D integrates into regions of a gene density equivalent to a random site in the human genome and showed that CPSF6 depletion causes WT HIV-1 to phenocopy N74D in this respect. We suggest that depletion of CPSF6 has a more profound effect upon integration site targeting than depletion of Nup358 or TNPO3 because it functions upstream of both of these cofactors and prevents the virus from interacting with either of them. We interpret these observations to mean that HIV-1 must use a CPSF6/Nup358/TNPO3-dependent nuclear entry pathway in order to integrate into sites of optimal gene density.

The field has taken to measuring the density of genes surrounding HIV-1 integration sites because almost all genomic features that correlate with HIV-1 integration – such as transcriptional activity, histone markers, CpG islands, DNase I hypersensitivity sites, and GC content – have been shown to strongly correlate with gene density (138). As discussed in section 1.2.9.4, the fact that so many different genomic characteristics and features correlate with each other makes it difficult to discern which actually impact the virus. For example, until it was discovered that IN-LEDGF/p75 interaction determines integration inside genes without affecting gene density (138, 352), it could have been hypothesised that integration inside genes was a result of integrating in gene-rich regions, or vice versa. In this study, we found that neither N74D mutation nor CPSF6 depletion prevented HIV-1 from preferentially integrating inside genes. This shows that IN-LEDGF/p75 interaction is not dependent upon CA-CPSF6 interaction.

Models of HIV-1 integration site selection are not yet able to predict the pattern observed *in vivo* and so it is thought that there are other unidentified genomic characteristics that influence the provirus (138, 163). For example, Lusic M. *et*

al (unpublished observations) used DNA FISH of HIV-1 LTRs to locate proviruses within the nucleus and discovered that HIV-1 integrates in close proximity to the nuclear envelope. The nuclear periphery is generally associated with densely packed heterochromatin, but it has recently been shown that NPCs are associated with euchromatin (353). Lusic M. *et al.* (unpublished observations) also found that N74D integrates much further away from the nuclear envelope than WT HIV-1, toward the centre of the nucleus. This tells us that nuclear architecture does not force the virus to integrate at the edge of the nucleus. It appears that HIV-1 has evolved to integrate close to the NPC, which suggests that integrating here must somehow be beneficial to the virus. One hypothesis is that HIV-1 has evolved to integrate as quickly as possible once inside the nucleus, in order to avoid recognition of its free DNA ends by host DNA damage sensors such as DNA-dependent protein kinase (DNA-PK) (354). Another hypothesis is that integrating near the NPC is beneficial to the virus because it results in integration into actively transcribed genes. This is supported by recent reports that nucleoporins specifically recruit transcription and RNA processing machinery in order to link mRNA synthesis and export (355, 356). Intriguingly, it has been proposed that NPCs within a cell are heterogeneous in composition and function (357). If so, CPSF6 may only traffic to NPCs at which active transcription and mRNA processing are occurring. It is therefore possible that HIV-1 evolved to interact with CPSF6 to only to be targeted to NPCs, but specifically to NPCs associated with active transcription. We hypothesise that HIV-1 integrates soon after traversing the NPC because it is fed into a tightly regulated nuclear entry pathway through interactions with a series of nuclear entry cofactors, including CPSF6, Nup358, and TNPO3. We propose that N74D is not fed into this pathway because it does not interact with this series of nuclear entry cofactors. Therefore, once inside the nucleus, it travels further before eventually integrating.

Given our findings that CA-CPSF6 interaction determines downstream interaction with Nup358, TNPO3, and Nup153 and that depletion of CPSF6, Nup358, or TNPO3 leads to integration into regions of a lower gene density than in control cells, we predicted that depletion of Nup153 would have a

similar effect. We were surprised to find that Nup153 depletion did not have a significant effect upon integration site characteristics of HIV-1. Whilst Di Nunzio et al. (84) reported similar findings to this study, Koh, et al. (268) reported that Nup153 depletion caused a very modest but statistically significant decrease in gene density when examining a very large number of integration sites. However, they reported that Nup153 depletion caused a similarly small decrease in gene density surrounding N74D proviruses, suggesting that this phenotype is not controlled by CPSF6. Given that Nup153 interacts with up to 25% of chromatin and its depletion alters expression of ~5,700 genes (262) (discussed in section 1.5.1.3), its depletion may also alter the positioning of DNA with respect to the NPC and therefore which regions HIV-1 integrates into.

CypA and CPSF6 play very similar roles in HIV-1 infection. Both interact directly with CA and determine downstream cofactor use. It is therefore surprising that this study and others have shown CypA and CPSF6 to have opposing effects upon HIV-1 integration site selection and nuclear positioning. CA-CPSF6 interaction increases the gene density surrounding proviruses whilst CA-CypA decreases the gene density surrounding proviruses (figure 25) (79). Furthermore, Lusic M. *et al.* (unpublished observations) found that CA-CPSF6 interaction increases the proximity of proviruses to the nuclear envelope whilst CA-CypA decreases the proximity of proviruses to the nuclear envelope. These observations reveal that, although the roles of CypA and CPSF6 in HIV-1 are linked, they are not redundant. Intriguingly, Schaller et al. (79) found that CsA treatment did not significantly affect the integration site preferences of N74D, indicating that a lack of CPSF6 is dominant over a lack of CypA with respect to HIV-1 integration site selection.

3.3.6 CPSF5 does not play a role in HIV-1 infection

In the nucleus, two monomers of CPSF6 interact via their RRMs with a dimer of CPSF5, forming the heterotetramer CFIm (315). CPSF5 is nuclear at steady state but shuttles in and out of nuclei efficiently when in complex with CPSF6 (317, 321, 322). It is therefore possible that cytoplasmic CPSF6 (which interacts with

HIV-1 CA) is in complex with CPSF5. We therefore tested whether CPSF5 is required for CPSF6 to interact with HIV-1 CA and to carry out its roles in HIV-1 infection. In this study, we have shown that GFP-HA-hCPSF6[68]$_{180\text{-}321}$ (which has no RRM and thus cannot interact with CPSF5) is able to restrict HIV-1, which implies that CPSF6 is able to interact with CA alone. Furthermore, we found that CPSF5 depletion does not affect HIV-1 infectivity, the sensitivity of HIV-1 to Nup358 or TNPO3 depletion, nor does it affect HIV-1 integration site selection. Together, these observations argue that CPSF5 does not influence HIV-1 infection or the role of endogenous CPSF6 in HIV-1 infection.

3.3.7 CypA and CPSF6 protects HIV-1 from innate immune detection

In the present study we have examined the role of CPSF6 in HIV-1 infection of HeLa cells. These cells are highly permissive and easily manipulated, which has allowed us to probe the effects of CPSF6 upon the early stages of HIV-1 infection. However, it was noted that, whilst WT HIV-1 replicates efficiently in MDMs, neither N74D nor P90A are able to replicate in MDMs (79, 92, 358) and so other studies focused upon the role of CPSF6 in these primary cells. Rasaiyaah *et al.* (92) demonstrated that, unlike WT HIV-1, N74D induces a type I IFN response in MDMs that is detectable by IFN ELISA, RT-qPCR of IFN-stimulated genes (ISGs), and microarray. They showed that N74D replication could be rescued by preventing the induction of an IFN response using an antibody that blocks the IFN-α/β receptor 2 (IFNAR2). In the absence of an IFN response, N74D replicated as efficiently as WT HIV-1, which confirmed that an antiviral response is the only cause of its defect in MDMs. Rasiyaah *et al.* went on to show that shRNA-mediated depletion of CPSF6 caused WT HIV-1 to phenocopy N74D and trigger an IFN response that prevented its own replication. Intriguingly, P90A mutation or treatment with CsA similarly caused HIV-1 to induce an inhibitory IFN response. For both N74D and P90A the induction of this antiviral response was shown to be dependent upon reverse transcription but not integration. In agreement with this, it has recently been shown that RNA-DNA hybrids are the pathogen-associated molecular patterns (PAMPs) that trigger cGAS (359). In agreement with this, the PRR of P90A

reverse transcripts was found to be cyclic GMP-AMP (cGAMP) synthase (cGAS), a recently identified cytosolic DNA sensor expressed in myeloid cells (360-362). However, the PRR of N74D in MDMs remains unknown. We interpret these data to mean that both CypA and CPSF6 are HIV-1 cofactors that prevent exposure of reverse transcripts to innate immune PRRs by interacting directly with HIV-1 cores. As discussed in sections 1.5.2.4 and 3.3.2, several studies have proposed that CypA and CPSF6 stabilise cores and therefore delay uncoating. Furthermore, several studies have now demonstrated that HIV-1 cores can be seen dock at NPCs (65, 100, 101) (see section 1.2.7.1). We therefore hypothesise that CypA and CPSF6 delay uncoating until immediately prior to nuclear entry, thereby minimising exposure of viral reverse transcripts to innate immune sensors.

The finding that N74D and P90A replicate as efficiently as WT HIV-1 in MDMs in the absence of an antiviral response suggests that the nuclear entry pathways taken by these viruses are as available and efficient in MDMs as they are in HeLa. We hypothesise that the reason N74D and P90A do not trigger an innate immune response in HeLa cells is because they have defective innate immune signaling pathways. For example, when cGAS detects cytoplasmic DNA (for example, during P90A HIV-1 infection of MDMs) it catalyses production of the second messenger cGAMP, which then activates the stimulator of IFN genes (STING) (363-365). However, it has been shown that HeLa cells are insensitive to addition of cGAMP but that the STING expressed by HeLa cells is functional in other cell types, suggesting that something downstream of STING may be defective in HeLa cells (Tan CP, unpublished observations). If would be of great interest to reconstitute this signaling pathway in HeLa cells and to test whether N74D and P90A induce an antiviral response as predicted.

3.3.8   A model of HIV-1 nuclear entry

The findings presented here and in the other studies discussed have lead us to a model of HIV-1 nuclear entry, which is depicted in figure 29. We propose that HIV-1 cores interact directly with cytoplasmic CypA and CPSF6, which stabilise
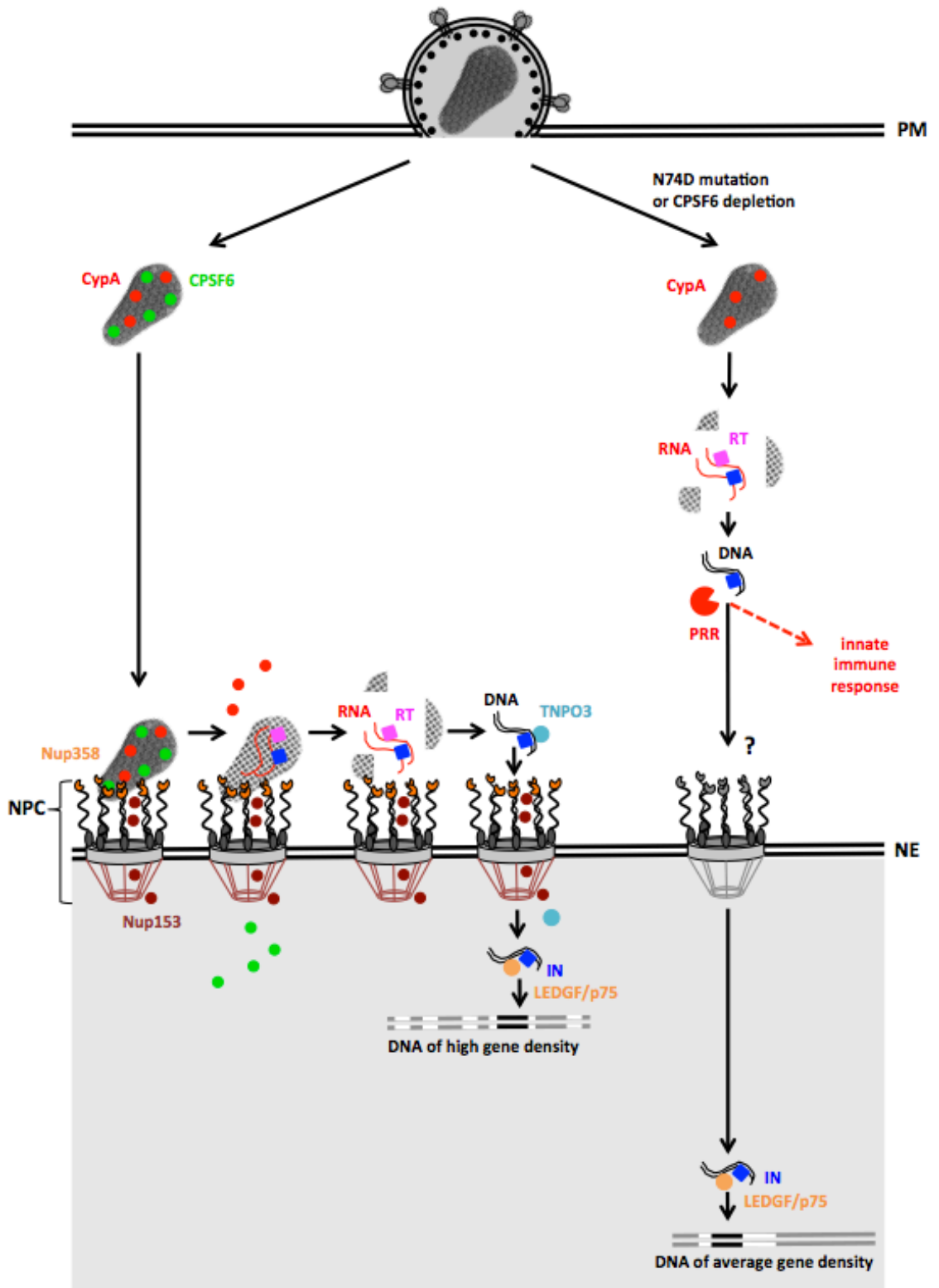
Figure 29. A model of HIV-1 nuclear entry
A cartoon model of our hypothesis of HIV-1 nuclear entry, with (left pathway) or without (right pathway) CA-CPSF6 interaction. PM = plasma membrane. NE = nuclear envelope. NPV = nuclear pore complex. Host cofactors are shown as circles. Viral proteins are shown as squares.

cores and prevent uncoating (and potentially reverse transcription). CPSF6 traffics to the nucleus due to its NLS/s and thus HIV-1 is also targeted to the nucleus. Having reached the NPC CypA and CPSF6 detach from viral cores, potentially due to competition from Nup358 and Nup153, respectively, and/or because CPSF6 enters the nucleus via its NLS. Once CypA and CPSF6 detach, the viral core is no longer stabilised and so it uncoats (and if reverse transcription were inhibited due to stabilisation, it would now be able to proceed). By uncoating in the vicinity of the NPC, the risk of viral DNA being detected by cGAS or other cytoplasmic PRRs is minimised. The viral PIC is now in an ideal location to interact with nuclear entry machinery such as TNPO3. Once inside the nucleus the viral cDNA integrates into the first LEDGF/p75-bound region of DNA it encounters, close to the edge of the nucleus. We hypothesise that this results in integration into the gene-rich and actively transcribed chromatin, which is associated with NPCs. If different types of NPC exist within a cell, we hypothesise that HIV-1 is targeted to NPCs at which active transcription is occurring because it was trafficked there by the mRNA processing factor CPSF6.

This model would mean that, in the absence of CPSF6, the viral core prematurely uncoats in the cytoplasm. In cells with intact innate DNA sensing pathways, such as primary MDMs, this would trigger an immune response and the virus would be unable to replicate (92). However, in cells such as HeLa, reverse transcription can proceed without inducing an IFN response and the viral DNA can access the nucleus through an alternative route. Although the details of this route are unknown, it is not dependent upon specific interaction with any of the nucleoporins or karyopherins examined in this study. However, it is dependent upon an intact NPC. Due to a lack of interaction with CPSF6 and the other components of the CPSF6-dependent nuclear entry pathway, the virus is not targeted to integrate into specific regions of DNA. It therefore integrates further inside the nucleus in regions of a gene density equivalent to random. However, the viral IN still interacts with LEDGF/p75 and so still integrates inside genes.

4 Chapter 4. CTL escape mutations in CA can alter the nuclear entry pathway of HIV-1

4.1 Introduction

4.1.1 The CTL response to viral infection

The majority of cytoplasmic and nuclear proteins – including viral proteins – are degraded by the proteasome into short peptides. A sample of these peptides are translocated into the ER lumen through a pore created by the transporter associated with antigen presentation (TAP) (366). In the ER, a peptide-loading complex assists in the assembling the MHC-I complex which, as shown in figure 30A, comprises a TAP-transported peptide, a transmembrane heavy chain, and β_2-microglobulin (367, 368). This MHC-I complex is then released from the ER and follows the secretory pathway onto the cell surface. Immature CTLs express TCRs that recognise peptides/antigens presented to them in the context of MHC-I. This TCR/MHC-1 complex is then bound by cluster of differentiation 8 (CD8), as shown in figure 30B, which prolongs and enhances the specificity of the interaction (369-371). TCRs only recognise foreign peptides (for example, viral peptides) because T cells that recognise 'self' peptides are negatively selected in the thymus. Unless an antigen is presented by a professional antigen presenting cell (APC), activation of CTLs requires a secondary signal from helper CD4+ cells. Activated CTLs produce IL-2, which drives their own clonal expansion. Mature CTLs then recognise other cells expressing this antigen and kill them through several different mechanisms. CTLs release perforin, which creates pores in cell membranes, and granzymes, which enter the cell through these pores and trigger apoptosis. CTLs also produce cytokines such as IFN-γ, tumour necrosis factor α (TNF-α), and chemokine ligands 3, 4, and 5 (CCL3, 4, and 5), which can induce an antiviral state in surrounding cells and so reduces the spread of infection (372).

4.1.2 The CTL response to HIV-1 infection

In acute HIV-1 infection, the appearance of an HIV-1 specific CTL response coincides with a rapid drop from peak viraemia. However, this CTL response is
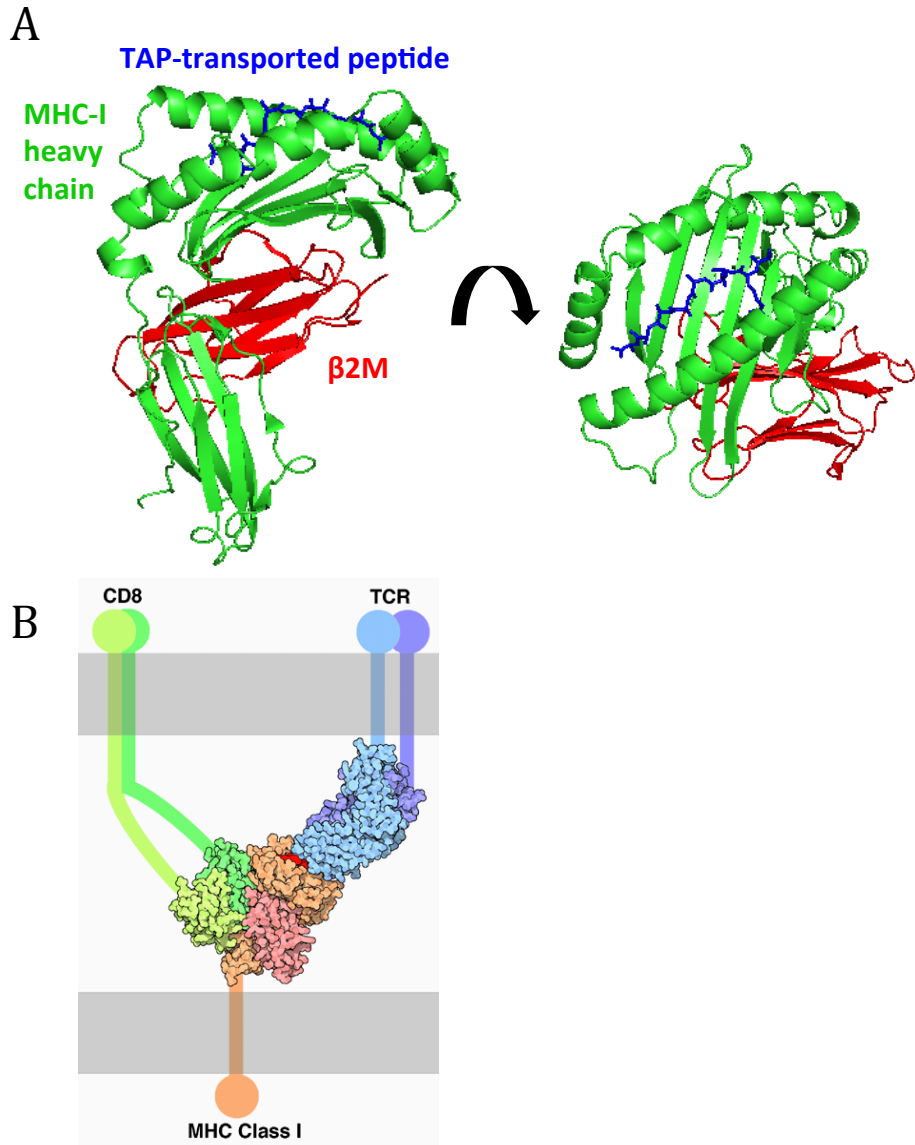
Figure 30. The MHC-I B*27 complex.
(A) A crystal structure of the MHC-I B*27 complex (PDB 1HSA from Madden et al. 1992). The MHC-I heavy chain is in green, β2-microglobulin in red, and the peptide 'ARAAAAAA' in blue, where 'R' is the anchor residue. **(B)** MHC-I in complex with the TCR and CD8 (Goodsell. D 2005, incorporates PDBs 1BD2 from Ding et al. 1998 and 1AKJ from Gao et al. 1997). CD8 dimer is in greens, TCR dimer in blue/purple, MHC-I heavy chain in orange, β2-microglobulin in pink, and the MHC-bound peptide in red. The grey bands represent plasma membranes, the upper of a CD8+ T cell and the lower of a cell present antigen to the CD8+ T cell in the context of MHC-I (potentially a virus-infected cell). The TCR-MHC-I interaction occurs first, and is stabilised by the interaction with CD8.

usually very narrow, targeting 1-3 epitopes usually in Env and/or Nef (373). In chronic infection CTL responses broaden; an average of 14 epitopes are targeted with the majority lying within Gag and Pol. Up to 19% of an infected individual's CTLs can be specific for HIV-1 and CTL levels remain high throughout infection until the onset of AIDS (374). Although CTLs are unable to eliminate HIV-1 infection, they are thought to be largely responsible for the ability of the immune system to control the virus for as long as it does. Two genome-wide association studies (GWASs) have sought to identify the major genetic determinants of the outcome of HIV-1 infection. They identified MHC class I alleles associated with both protection from and susceptibility to disease. In fact, polymorphisms in the MHC class I region of the genome account for 15-20% of the variance in HIV-1 viral load set point (375, 376). Furthermore, the importance of CTLs in SIVmac infection of Rhesus macaques has been demonstrated experimentally. Transient depletion of CTLs during acute infection lead to an irreversible loss of viral control, whilst transient depletion during chronic infection lead to a temporary loss of viral control (377).

Despite CTL responses being able to control the virus, they are not sufficient to clear the virus. Several reasons for the inability of CTLs to clear HIV-1 have been put forward. For example, the HIV-1 accessory protein Nef is produced early after infection and downregulates surface expression of MHC-I (378). Nef induces endocytosis of MHC-I molecules from the surface of infected cells, prevents them from being recycled to surface, and also diverts newly synthesized MHC-I to a paranuclear compartment (229). Nef expression has been shown to significantly reduce the ability of CTLs to kill HIV-1 infected cells (379). A few patients infected with Nef-deleted viruses have been identified and, although they are/were all LTNPs, they did not completely clear the virus (229). This suggests that whilst Nef does reduce the efficacy of CTL responses, it is not the only reason that CTLs cannot clear HIV-1. The main reason that CTL responses are unable to clear HIV-1 infection is thought to be that the virus can continually mutate the epitopes targeted by CTLs, thereby evading recognition. It had been hypothesised that the latent reservoir is seeded very early after infection and so, if the reservoir could be reactivated, CTLs would be able to

eliminate it. However, a recent study found that the reservoir is in fact dynamic and is constantly being reseeded, such that the vast majority of latent proviruses contain CTL escape mutations in the most commonly targeted epitopes and so could not be eliminated by CTLs (380).

4.1.3 Protective HLA alleles target epitopes in Gag

The human leukocyte antigen (HLA) genes are human MHC genes, and encode the highly variable heavy chain of the HLA complex. HLA-A, B, and C are highly expressed 'major' HLA genes and are the most polymorphic human genes with over 4,200 different alleles known so far (381). As shown in figure 30A, this variable heavy chain forms a groove that binds to the peptide being presented (382). This groove contains 57 residues that can potentially interact with a bound peptide, which are extremely variable between different HLA genes and alleles (due to strong positive selection from a huge number of pathogens). As a result, there is a huge diversity in the peptides that can be presented by HLA molecules within an individual and within a population (381).

Single genome analysis (SGA) has revealed that CTL escape mutations can be detected in HIV-1 as early as 25 days after infection (383), reflecting both the strong selective pressure put upon the virus by CTLs and the speed of HIV-1 evolution. Mutations within the HLA-restricted epitope either prevent the peptide-MHC interaction or recognition of this complex by the TCR. Mutations flanking the epitope can also impair peptide processing and presentation (384). As well as allowing escape from the CTL response within an individual, CTL escape mutations can accumulate in a population over time. For example, B*51:01 was protective for clade B infections early in the HIV-1 epidemic, but the accumulation of escape mutations within B*51:01-restricted epitopes in the population means it is no longer associated with protection (381).

It is not clear why some alleles are more protective than others, but many suggestions have been made. For example, it has been proposed that protective HLA alleles cause higher levels of CTL proliferation (381) and are more cross-

reactive (385). A major factor in determining the level of protection provided by a given allele appears to be the epitope it targets. The alleles most strongly associated with low viral load are B*57, 58, 81, 14, and 27, which, intriguingly, all target epitopes in Gag, mostly in the CA-encoding region (374). Several studies have now demonstrated that the breadth of Gag-specific CTL responses correlates with $CD4^+$ T cell count and negatively correlates with viral load (386-388). There are two main reasons why targeting Gag (and specifically CA) is thought to be more protective than other regions of the genome. Firstly, it has been shown that presentation of Gag-derived peptides (and CTL-mediated killing of these cells) occurs as early as 2 hours post-infection and is not dependent upon integration (389). This suggests that Gag-derived proteins (MA, CA, NC, and p6) are abundant enough in incoming virions to be processed and their peptides presented by MHC-I without *de novo* protein synthesis. This is in contrast to peptides derived from Env, for example, which are dependent upon *de novo* protein synthesis and are therefore not presented by MHC-I until much later in infection. Rapid presentation of viral peptides allows rapid recognition and killing of infected cells by CTLs, before Nef is produced and is able to downregulate MHC-I-mediated antigen presentation (see section 1.4.1) (389). The second reason is that Gag – and particularly CA – is one of the most highly conserved regions of the viral genome and so CTL escape mutations frequently cause significant fitness costs to the virus (390, 391). Viral fitness can sometimes be restored through compensatory mutations. However, the appearance of these compensatory mutations is often slow due to the low replicative capacity (and thus evolution) of the escaping virus and the fact that multiple mutations are often required for full compensation (325, 392).

4.1.4 HLA B*57 and B*27 CTL escape mutations in CA have infectivity defects that can be rescued by CsA

HLA alleles B*57 and B*5801 are associated with HIV-1 control in acute infection and slow disease progression (393). Both $B*57^+$ and $B*5801^+$ patients mount an immunodominant CTL response against the epitope TW10 in CA ($_{108}$TSTLQEQIGW$_{117}$) (394). In 84% of $B*57^+$ patients and 63% of $B*5801^+$

patients infected with clade C viruses (similar values are seen for clade B viruses) the virus escapes this CTL response by evolving the mutation T110N in CA, a mutation that has never been detected in patients without these HLA types. In patients infected with clade B viruses HLA-B*57 is also associated with the mutation G116A (395). Like many CA mutations, T110N causes a replication defect (325) and the fitness cost of this mutation *in vivo* is reflected in its quick reversion after transmission to HLA mismatched hosts (395). T110N is frequently found in the context of compensatory mutations such as I91V, M96I, and H87Q. Intriguingly, these compensatory mutations lie within/adjacent to the CypA binding loop of CA and this flexible loop lies in close proximity to TW10 in secondary structure, as shown in figure 31A. It is therefore unsurprising that these compensatory mutations have been shown to confer resistance to CsA in Jurkat cells (325).

Another protective HLA allele is HLA B*27, which targets an epitope in CA called KK10 ($_{131}$KRWIILGLNK$_{140}$) (393). Some HLA B*27$^+$ patients never develop any CTL escape mutations and remain LNTPs. The majority of HLA B*27$^+$ patients develop the mutation L136M within a few years of infection (396). Although L136 lies within KK10, it does not interact directly with HLA but rather with the TCR (397, 398). L136M is thought to escape the CTL response, but the epitope remains sufficiently immunogenic to illicit another response very quickly afterward and so the virus remains controlled (399). In a minority of HLA B*27$^+$ patients the escape mutation R132K (or rarely R132Q/G/T (400)) arises. R132K mutation escapes the CTL response by reducing the affinity of KK10 for HLA by 348-fold (401). The appearance of R132K usually arises in the context of L136M, arises late in infection (on average 9-12 years post-infection), and is associated with progression to AIDS (381, 396). Schneidewind *et al.* found that whilst L136M mutation does not affect HIV-1 replication in Jurkat cells, R132K mutation causes a dramatic replication defect (326). They therefore hypothesised that R132K must be accompanied by a compensatory mutation within CA that restores viral fitness. Indeed, they found that R132K is almost always accompanied by S41A, which completely restores replication in Jurkat cells. We hypothesise that R132K
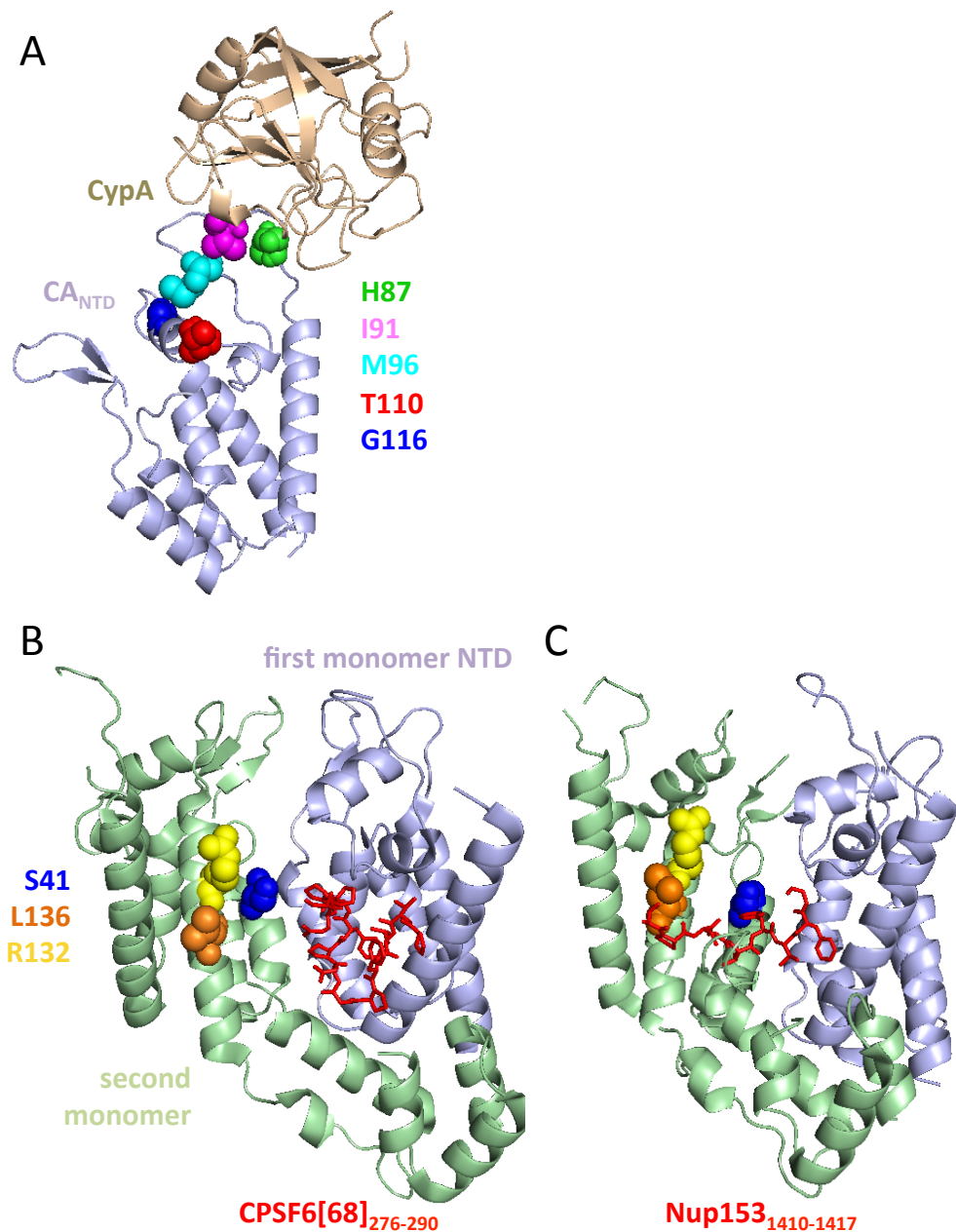
Figure 31. The location of HLA-B*57 and B*27-associated mutations in CA
(A) A crystal structure of CypA in beige bound to CA_{NTD} in lilac (PDB 1AK4, Gamble et al. 1996). B*57 associated residues are highlighted with coloured spheres. The CypA binding loop of CA is defined as $_{85}PV\underline{H}AGP\underline{I}AP_{93}$. H87 in green and I91 in pink therefore lie within the CypA binding loop, whereas M96 in cyan, T110N in red, and G116 in blue lie proximal to it. **(B)** Crystal structure of $CPSF6[68]_{276-290}$ in complex with a dimer of hexameric CA (PDB 4U0A from Price et al. 2014). **(C)** Crystal structure of $Nup153_{1410-1417}$ in complex with a dimer of hexameric CA (PDB 4U0C from Price et al. 2014). **(B-C)** First CA monomer NTD in lilac. Second CA monomer NTD and CTD in green. CPSF6/Nup153 peptides in red. B*27-associated residues highlighted with coloured spheres with L136 in orange, R132 in yellow, and S41 in blue.

mutation may arise quite quickly but that it replicates so poorly that it does not grow out and so is not detected by sequencing. Furthermore, due to the very slow replication of R132K, it takes many years for the compensatory mutation S41A to arise.

Like the B*57-associated mutant T110N, the B*27-associated mutant R132K has an altered relationship with CypA. R132K mutation reduces the replicative capacity of HIV-1 by over 100-fold in Jurkat cells, but its titre is rescued to the level of WT HIV-1 by the addition of CsA or in CypA$^{-/-}$ cells (326). L136M mutation does not appear to affect the titre of the virus or its relationship with CsA. However, the compensatory mutation S41A restores the WT phenotype to R132K-containing viruses. Although S41 does not lie within the KK10 epitope in helix 7, it lies in helix 2, meaning it is adjacent to KK10 in the CA$_{NTD}$. We have noted that these three B*27–associated residues lie in/near the pocket of CA bound by CPSF6 and Nup153 as shown in figures 31B and C, respectively. We therefore hypothesise that, as well as having an altered relationship with CypA, these mutations may affect interaction with CPSF6.

As discussed, mounting evidence suggests that the ability of some CTL responses to control infection better than others can be at least partially attributed to the fitness costs associated with escaping these responses. The fact that two of the HLA types most strongly associated with viral control – B*57 and B*27 – put pressure upon the virus to develop mutations that alter the relationship of the virus with CypA, illustrates the importance of CypA as an HIV-1 cofactor *in vivo*. Given our interest in CypA, particularly its relationship with CPSF6, and the effects of these two cofactors upon HIV-1 nuclear entry and integration site targeting, we sought to test whether CTL escape mutations that alter the relationship of the virus with CypA also differ in their relationship with other components of the nuclear entry pathway.

4.2 Results

4.2.1 CA mutations associated with escape from HLA-B*57-restricted CTLs have an altered relationship with CypA but interact normally with other members of the nuclear entry pathway

Brockman, et al. (325) reported that the mutation T110N in CA, which is strongly associated with escape from B*57-restricted CTLs, conferred a significant delay in replication in PBMCs and Jurkat T cells. They also reported that T110N increased the sensitivity of the virus to inhibition by CsA, but that the virus could be made insensitive to CsA by co-mutation of 4 other HLA-B*57-associated mutations (H87Q, I91V, M96I, and G116A). Having found that these mutations are all located within or in close proximity to the CypA-binding loop (figure 31A), we sought to test the effects of each mutation individually upon viral titre and CsA sensitivity. In a single-round infection of HeLa cells we found that T110N was indeed 7-fold less infectious than WT HIV-1 when normalised to ngRT, but that H87Q, I91V, M96I, and G116A also caused a 3 to 5-fold reduction in viral titre (figure 32A). In the presence of 5μM CsA, T110N and G116A were inhibited 3- and 2–fold, respectively, whereas neither WT HIV-1 nor any of the other HLA-B*57-associated mutants were significantly affected by CsA in HeLa cells (figure 32B).

As discussed in section 1.5.2.5, CypA has been shown to affect the interaction of HIV-1 with downstream nuclear entry cofactors (79) and so we sought to test whether other aspects of the nuclear entry pathway were affected by these mutations. In order to test the ability of these virus to infect non-dividing cells we pre-treated HeLa cells with ApC 24-hours prior to infection; ApC is a fungal metabolite that specifically inhibits DNA polymerase and so arrests cell division in early S phase. We found that ApC did not significantly affect any of the viruses tested (figure 32C), suggesting that all viruses can infect via the NPC.

Data from our lab suggest that altered interaction with CypA – due to either CsA treatment or P90A mutation – reduces the ability of HIV-1 CA to interact with CPSF6 (Hilditch. L, unpublished observations). We therefore also tested
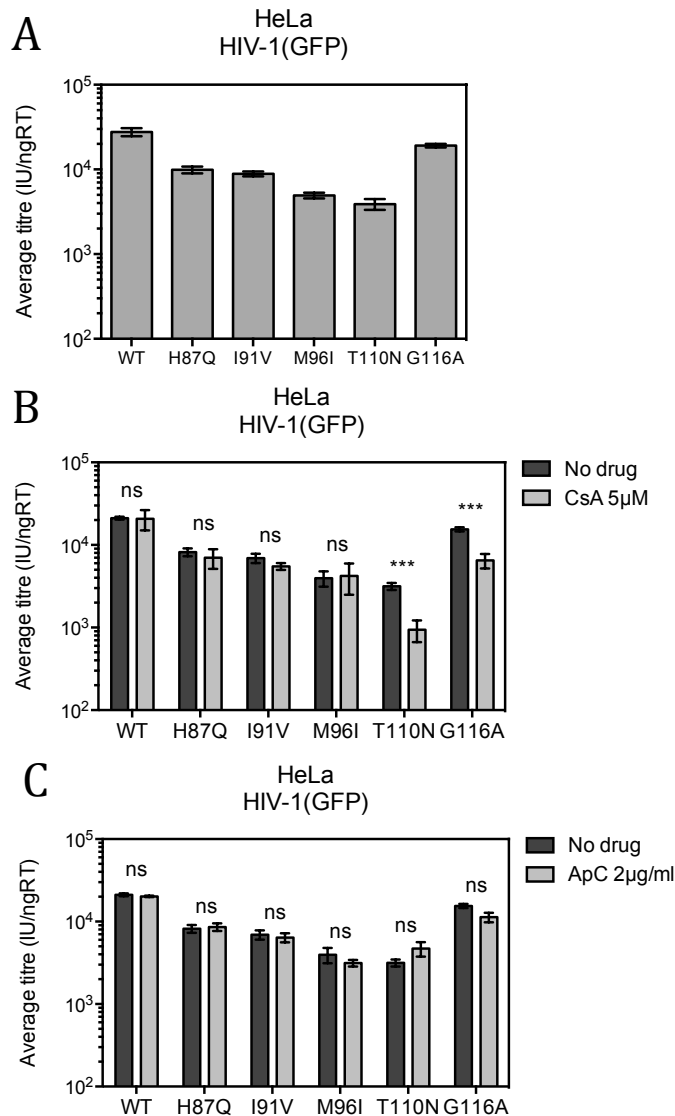
Figure 32. B*57-associated mutations alter the relationship of HIV-1 with CsA but not ApC

HeLa cells that were **(A)** unmodified, **(B)** +/- 5µM CsA, **(C)** +/- ApC, were infected with a titration of HIV-1(GFP) vectors carrying B*57-associated mutations and viral titre was determined 48hrs p.i. by flow cytometry. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean). ns means $p>0.05$ (not statistically significant, * means $p\leq0.05$, ** means $p\leq0.01$, *** means $p\leq0.001$, as measured by Student's *t*-test.

whether the increased sensitivity of T110N or G116A to CsA impacted the sensitivity of these mutants to inhibition by HA-hCPSF6[68]-358, and included the CPSF6 binding mutant N74D as a negative control. We found that all mutants were inhibited to a similar degree by this truncated CPSF6 construct (raw/normalised data in figure 33A/B). It has also been previously shown that CA-CypA interaction affects the sensitivity of HIV-1 to depletion of the nuclear entry cofactors TNPO3 or Nup358 (79) and so we tested the effects of RNAi-mediated depletion of TNPO3 or Nup358 upon the infectivity of the HLA-B*57-associated mutants, again using N74D as a negative control. Whereas WT, H87Q, and G116A were all significantly inhibited 4 to 5-fold by Nup358 depletion, I91V, M96I, and T110N were inhibited around 2-fold, a difference that was not statistically significant (figure 33C). In contrast, we found that all viruses tested were similarly sensitive to TNPO3 depletion. Taken together, these data suggest that the increased sensitivity of T110N and G116A to CsA does not result in any major changes in the way these viruses interact with downstream cofactors such as CPSF6, Nup358, and TNPO3.

4.2.2 The HLA-B*27 escape mutation R132K in CA causes a significant post-reverse transcription pre-nuclear entry defect in cell lines and primary CD4$^+$ T cells, which is rescued by the compensatory mutation S41A

As discussed in section 4.1.4, HLA B*27 is associated with the mutations L136M, R132K, and S41A in CA. Whilst L136M is reported not to affect HIV-1 titre in Jurkat or CEM cells, R132K causes a significant infectivity defect, which is compensated for by S41A (326, 400). Firstly, we tested the titre of these mutations – both individually and in the combinations in which they arise – in HeLa cells (figure 34A). In a single round infection, we found L136M to have no significant effect upon viral infectivity. As expected, R132K caused a 10 to 11-fold defect, regardless of whether position 136 was leucine or methionine. S41A mutation did not affect viral titre alone, but it completely rescued the titre of R132K or R132K/L136M. We subsequently focused on the combinations of mutations that arise *in vivo* in the order in they arise: L136M, R132K/L136M,
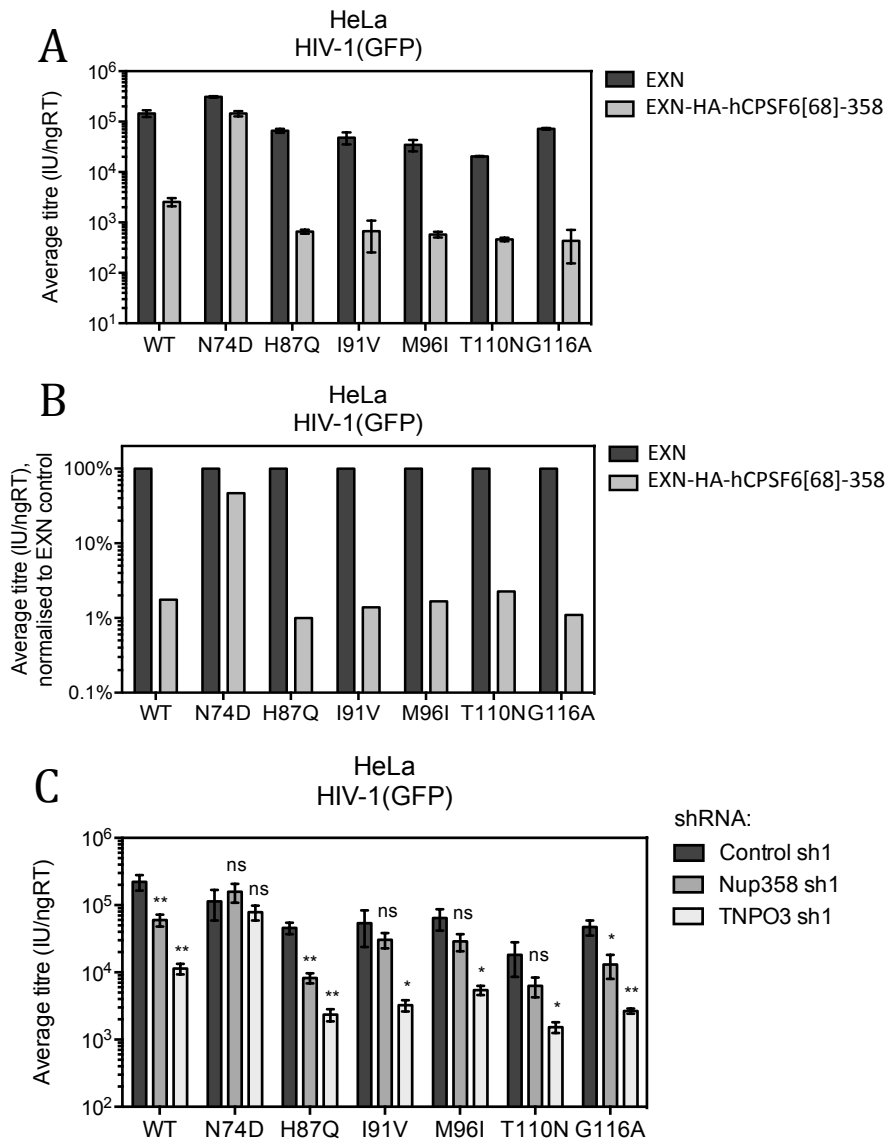
Figure 33. B*57-associated mutations do not alter the relationship of HIV-1 with hCPSF6[68]-358, Nup358, or TNPO3

HeLa cells stably expressing **(A/B)** hCPSF6[68]-358 or empty vector or **(C)** Nup358- or TNPO3-targeting shRNA or a non-targeting control shRNA, were infected with a titration of HIV-1(GFP) vectors carrying B*57-associated mutations and viral titre was determined 48hrs p.i. by flow cytometry. (B) shows data in (A) normalised to viral titre in control cells. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean). ns means p>0.05 (not statistically significant, * means p≤0.05, ** means p≤0.01, *** means p≤0.001, as measured by Student's *t*-test.
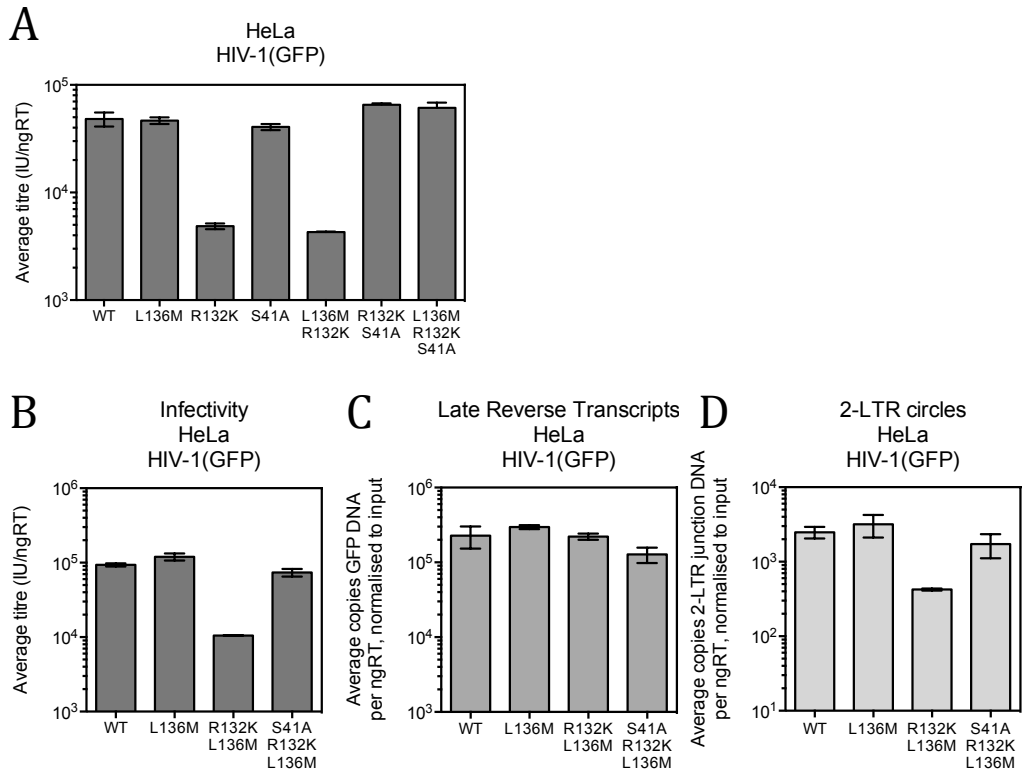
Figure 34. HLA-B*27-associated CTL escape mutant R132K causes a defect in HIV-1 reverse transcription, which is rescued by the compensatory mutation S41A

(A) HeLa cells were infected with a titration of HIV-1(GFP) vectors containing B*27-associated mutations and viral titre was measured 48hrs p.i. by flow cytometry. **(B-D)** HeLa cells were infected with a titration of WT, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP). **(B)** Viral titre was measured 48hrs p.i. by flow cytometry. **(C)** Late reverse transcripts were measured 6hrs p.i. by qPCR. **(D)** 2-LTR circles were measured 6hrs p.i. by qPCR. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

and S41A/R132K/L136M. In disagreement with a previous report that R132K/L136M was compromised for reverse transcription (326), we found that R132K/L136M formed late reverse transcripts as efficiently as WT HIV-1 (figure 34C). Instead, we found it was defective in forming 2-LTR circles (figure 34D), suggesting that this mutant is impaired prior to/at the level of nuclear entry.

Due to previous reports that this phenotype was cell-type dependent, we titrated these three viruses onto cells frequently used in HIV-1 assays – HeLa, TE671, Jurkat, CEM, and THP-1 - as well as CRFK cells (which are used in section 4.2.3) (figure 35A). R132K/L136M exhibited a significant infectivity defect in all cell lines, although the magnitude ranged from 3-fold in TE671s to over 250-fold in CEMs. Importantly, this mutant similarly exhibited a 5-fold infectivity defect in primary activated CD4$^+$ T cells (figure 35B), the cell type HIV-1 infects most commonly *in vivo*.

4.2.3 Infectivity of R132K/L136M is rescued to WT levels by CsA and R132K/L136M reduces the sensitivity of HIV-1 to omTRIMhuCypA

Schneidewind, et al. (326) reported that treatment with 5μM CsA or CypA knock-out in Jurkat cells inhibited WT HIV-1 but rescued R132K, such that both viruses were equally infectious. We were able to recapitulate this observation in Jurkat cells and TE671s. However, in HeLa, CEM, and THP-1 cells, we found that CsA did not significantly inhibit WT HIV-1 and that it completely rescued R132K/L136M (figure 36A). L136M and S41A/R132K/L136M both exhibited the WT phenotype with respect to infectivity and CsA in all cell types tested. This phenotype of cell-type dependent CsA resistance/dependence was originally observed for the CA mutant A92E, which evolves when HIV-1 is serially passaged in the presence of CsA (297, 402). Unlike R132K, A92E has never been identified *in vivo*. As discussed in section 1.5.2.3, this A92E phenotype is not understood but some have attributed it to an elusive CypA-dependent restriction factor. We were interested to compare the phenotypes of
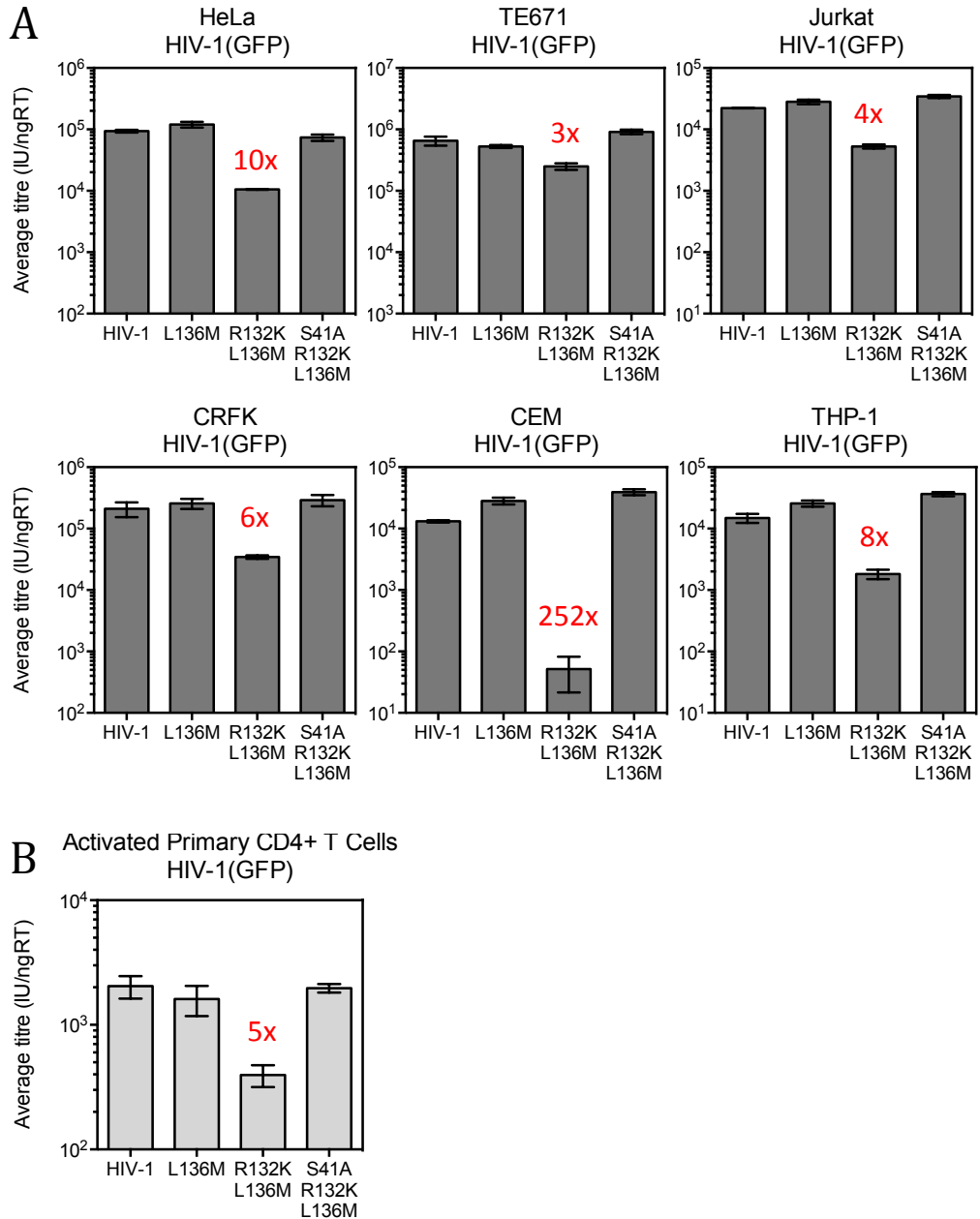
Figure 35. R132K/L136M is defective in a range of cell lines and in activated primary CD4+ T cells

(A) HeLa, TE671, Jurkat, CRFK, CEM, or THP-1 cells or **(B)** activated primary CD4+ T cells were infected with a titration of WT, L136M, R132K/L136M, or S41A/ R132K/L136M HIV-1(GFP) and viral titre measured 48hrs p.i. by flow cytometry. Red text shows fold decrease in titre caused by R132K in the context of L136M. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

182
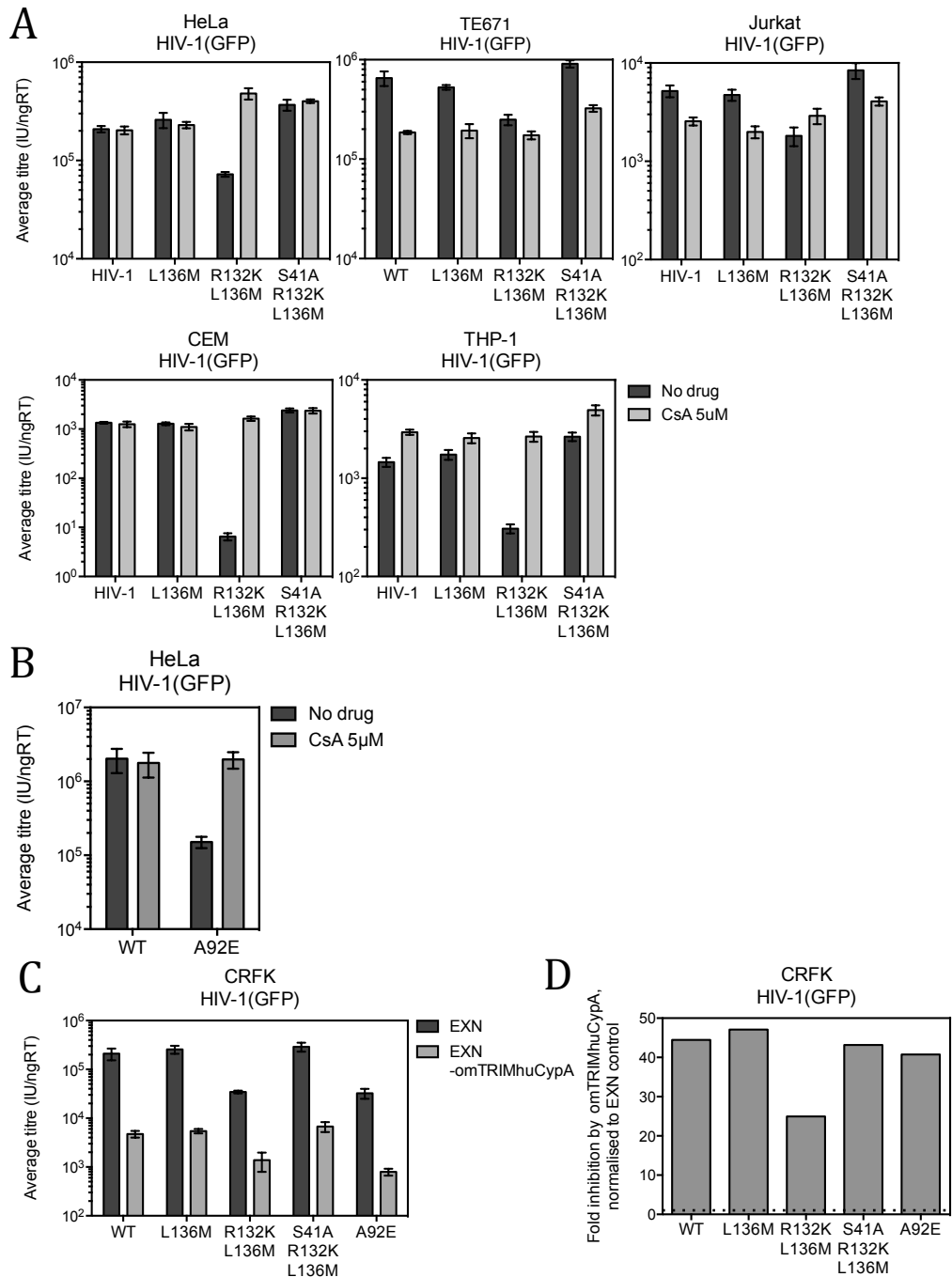
Figure 36. R132K or A92E mutation in CA alters the relationship between HIV-1 and CypA

(A) HeLa, TE671, Jurkat, CRFK, CEM, or THP-1 cells were infected with a titration of WT, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP) with or without 5μM CsA and viral titre was measured 48hrs p.i. by flow cytometry. **(B)** HeLa cells were infected with a titration of WT or A92E HIV-1(GFP) with or without 5μM CsA and viral titre was measured 48hrs p.i. by flow cytometry. **(C)** CRFK cells stably expressing omTRIMhuCypA or empty vector were infected with a titration of WT, A92E, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP) and viral titre was measured 48hrs p.i. by flow cytometry. **(D)** shows the data in (C) as fold difference between titre in control cells and titre in omTRIMhuCypA-expressing cells. The dotted line indicates where fold change = 1. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

183

R132K/L136M and A92E and were able to recapitulate the CsA-dependent phenotype of A92E in HeLa cells (figure 36B).

The finding that CsA is beneficial for R132K/L136M and A92E indicates that CypA binding to these mutated cores has negative effects upon viral infectivity. We therefore sought to test whether these mutations alter the interaction between CA and CypA. As a surrogate measurement of the affinity of CA for CypA, we measured restriction by omTRIMhuCypA (provided by Schaller. T), an artificial construct in which the cyclophilin domain of the naturally occurring HIV-1 restriction factor omTRIMCyp (see section 1.3.1 for details of restriction mechanism) is substituted with human CypA. The ability of this protein to restrict the virus is determined by the ability of CypA to bind to CA. We expressed this construct in CRFK cells because these feline cells do not express TRIM5, which can form non-functional dimers with other TRIM constructs. We found that omTRIMhuCypA restricts WT, L136M, S41A/R132K/L136M, and A92E HIV-1 by 40 to 47-fold. In contrast, it only restricts R132K/L136M by 25-fold (raw/fold data in figure 36C/D). This suggests that R132K mutation reduces the ability of CA to interact with CypA but that the compensatory mutation S41A restores the affinity to that of WT CA. This observation is particularly interesting because it suggests that, despite exhibiting the same CsA-dependent/resistant phenotype, R132K/L136M and A92E differ in their ability to bind to CypA.

4.2.4 R132K/L136M is inhibited by Aphidicolin, but S41A mutation or CsA rescues the virus from this

Whereas WT HIV-1 infects non-dividing cells efficiently, A92E infectivity is reportedly reduced by cell-cycle arrest (298, 403). We were able to recapitulate this observation in this study (figure 37A). We were therefore interested to test the effect of ApC-induced cell-cycle arrest upon the HLA-B*27-associated mutants. As shown in figure 37B, R132K/L136M was inhibited 7-fold by ApC in HeLa cells, whilst WT, L136M, and S41A/R132K/L136M HIV-1 were all insensitive to ApC. In TE671, Jurkat, and THP-1 cells, WT, L136M, and
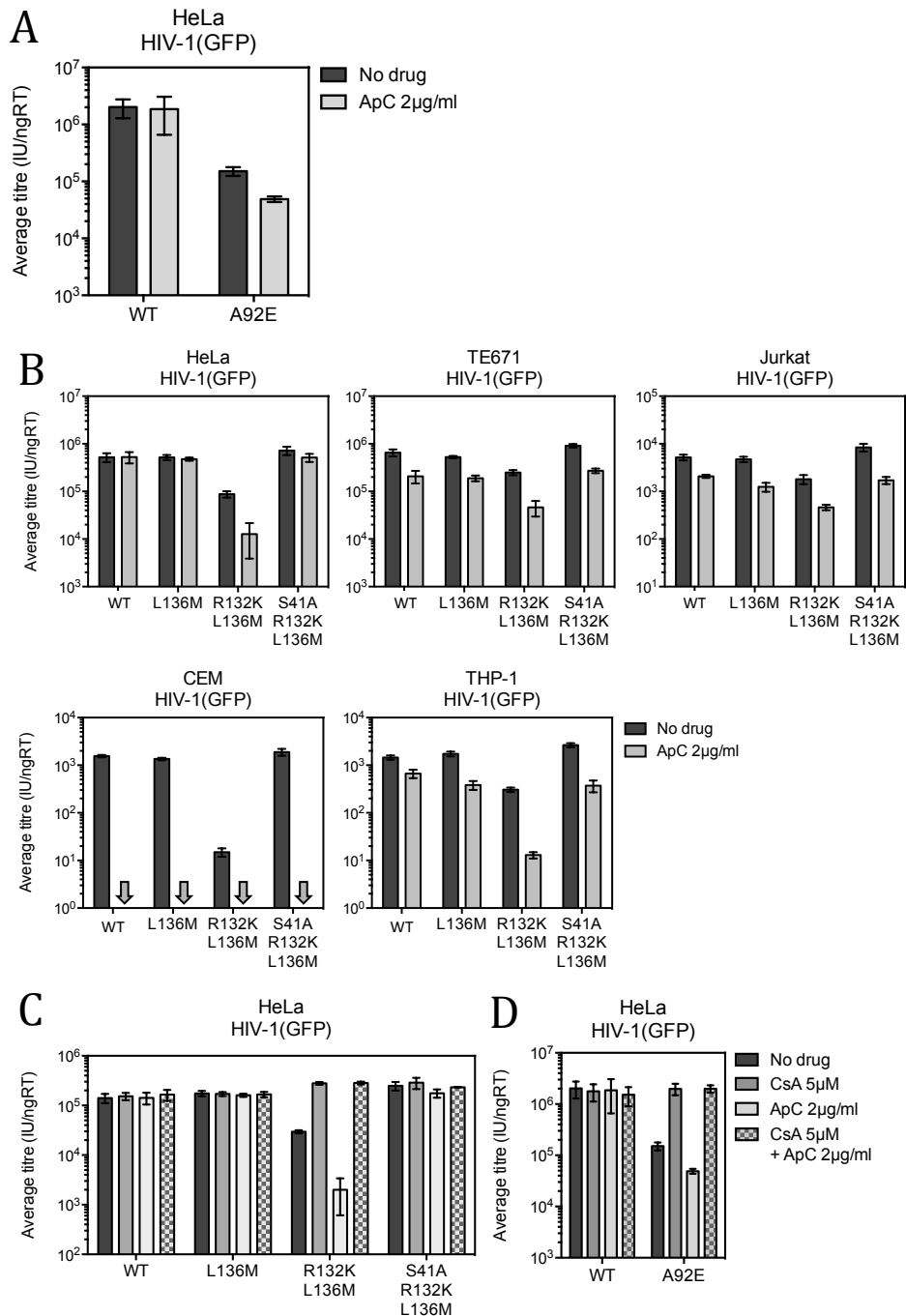
Figure 37. R132K or A92E mutation in CA causes HIV-1 to become cell-cycle dependent

(A) HeLa cells were infected with a titration of WT or A92E HIV-1(GFP) with or without 2μg/ml ApC and viral titre was measured 48hrs p.i. by flow cytometry. **(B)** HeLa, TE671, Jurkat, CEM, or THP-1 cells were infected with a titration of WT, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP) with or without 2μg/ml ApC and viral titre was measured 48hrs p.i. by flow cytometry. Grey arrows indicate values below the level of detection. **(C)** HeLa cells were infected with a titration of WT, L136M, R132K/L136M, or S41A/R132K/L136M with 5μM CsA, 2μg/ml ApC, both CsA and ApC, or neither drug, and viral titre was measured 48hrs p.i. by flow cytometry. **(D)** As in (C) but with WT or A92E HIV-1(GFP). All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

185

S41A/R132K/L136M HIV-1 were modestly inhibited by ApC (2 to 3-fold) but R132K/L136M was inhibited more so, from 4 to 24-fold. We have previously shown that R132K mutation reduces the efficiency with which HIV-1 can enter the nucleus (figure 34D). The fact that the infectivity defect of R132K/L136M is exacerbated in non-dividing cells suggests that the virus is inefficient at entering the nucleus via the NPC and that, in dividing cells, this mutant may enter the nucleus during mitosis. In CEM cells ApC inhibited all viruses tested, including WT HIV-1, to such an extent that any infection was below the limit of detection of this assay. The inability of HIV-1 to infect non-dividing CEM cells has previously been reported, but remains unexplained (404).

Given that CsA rescues the nuclear entry defect of R132K/L136M in unmodified HeLa cells, we hypothesised that CsA would also rescue its nuclear entry defect in arrested HeLa cells. Indeed, in the presence of CsA, ApC had no effect upon the infectivity of R132K/L136M or A92E (figure 37C and D, respectively). These observations suggest that the reduced ability of R132K/L136M and A92E to enter the nucleus via the NPC is due (either directly or indirectly) to the interaction of CA with CypA.

4.2.5 R132K reduces the sensitivity of HIV-1 to Nup358, TNPO3, or Nup153 depletion and to omTRIMhuN358Cyp-mediated restriction, but S41A restores the WT phenotype

Having found that A92E and R132K/L136M are compromised for nuclear entry and that R132K/L136M has a reduced interaction with CypA, we sought to test whether these mutations altered the dependence of HIV-1 upon the other known HIV-1 nuclear entry cofactors. Whereas depletion of Nup358 or TNPO3 inhibited WT HIV-1 7- or 22-fold, respectively, these depletions only inhibited R132K/L136M 2- or 4-fold (figure 38A). The compensatory mutation S41A largely restored the dependence of this virus upon these cofactors; S41A/R132K/L136M was inhibited 5- or 13-fold by Nup358 or TNPO3 depletion, respectively. A similar pattern was observed with transient depletion of Nup153, which inhibited WT, L136M, and S41A/R132K/L136M HIV-1
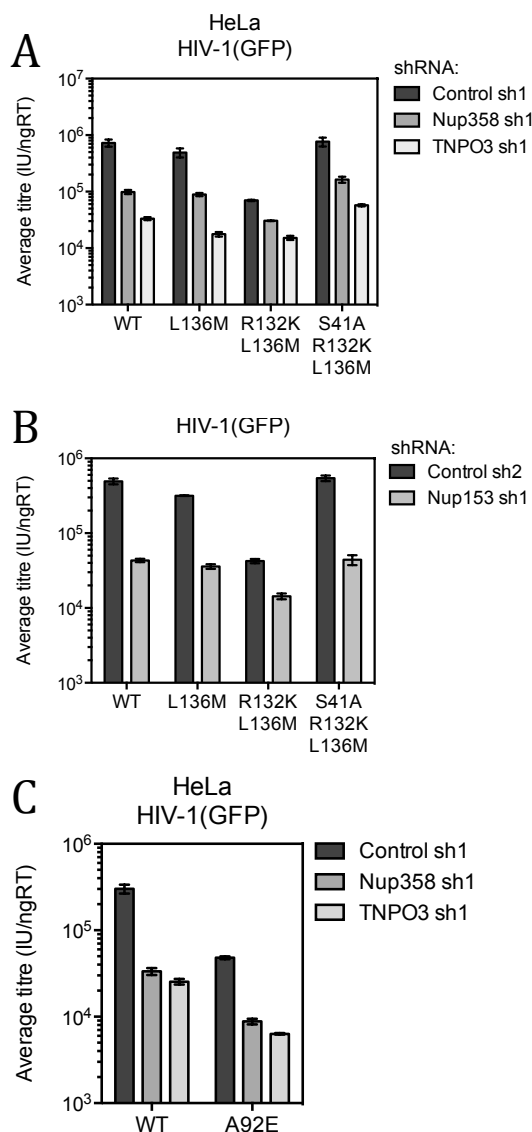
Figure 38. R132K, but not A92E, mutation in CA causes HIV-1 to become less dependent upon nuclear entry cofactors and S41A co-mutation restores sensitivity

(A) HeLa cell clones stably expressing Nup358- or TNPO3-targeting shRNA or non-targeting control shRNA were infected with a titration of WT, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP) and viral titre measured 48hrs p.i. by flow cytometry. **(B)** HeLa cells transiently expressing Nup153-targeting shRNA or non-targeting control shRNA were infected with a titration of WT, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP) and viral titre measured 48hrs p.i. by flow cytometry. **(C)** As in (A) but with WT or A92E HIV-1(GFP). All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

between 9 and 12-fold but only inhibited R132K/L136M 3-fold (figure 38B). These data suggest that R132K/L136M does not, and perhaps cannot, utilise the normal nuclear entry pathway of HIV-1. In contrast, A92E mutation did not affect the sensitivity of HIV-1 to Nup358 or TNPO3 depletion (figure 38C). This suggests that the reason R132K/L136M is less sensitive to depletion of Nup358 and TNPO3 is not necessarily related to its CsA/ApC phenotypes.

It has previously been reported that CsA causes WT HIV-1 to become insensitive to Nup358 depletion and less sensitive to TNPO3 depletion (79). We were able to recapitulate this observation and subsequently tested whether CsA was also able to determine the nuclear entry pathway of our three B*27-associated mutants. In the presence of 5μM CsA all viruses tested were of similar titres, were insensitive to Nup358 depletion, and were inhibited approximately 5-fold by TNPO3 depletion (figure 39A). This means that CsA caused R132K/L136M to gain dependence upon TNPO3, showing that the reduced interaction between R132K/L136M CA and CypA cannot be the cause of its reduced dependence upon TNPO3.

As discussed in section 1.5.1.2, $Nup358_{Cyp}$ interacts directly with the CypA binding loop of CA (79). Having found that R132K mutation reduces the sensitivity of HIV-1 to Nup358 depletion and to omTRIMhuCypA-mediated restriction, we hypothesised that this mutation may also reduce the affinity of CA for $Nup358_{Cyp}$. As a surrogate for $Nup358_{Cyp}$ binding, we measured the level of restriction caused by omTRIMhuN358Cyp (provided by Schaller. T), an artificial construct in which the cyclophilin domain of omTRIMCyp is substituted with $Nup358_{Cyp}$. We observed a similar pattern to that seen with omTRIMhuCypA; WT, L136M, S41A/R132K/L136M, and A92E HIV-1 were restricted 25 to 32-fold, whereas R132K/L136M was only restricted 16-fold (raw/fold data in figure 39B/C). Therefore, R132K mutation appears to reduce the interaction of CA with both CypA and $Nup358_{Cyp.}$ However, it remains unclear whether the insensitivity of R132K/L136M to Nup358 depletion is due to its reduced interaction with Nup358 or not.
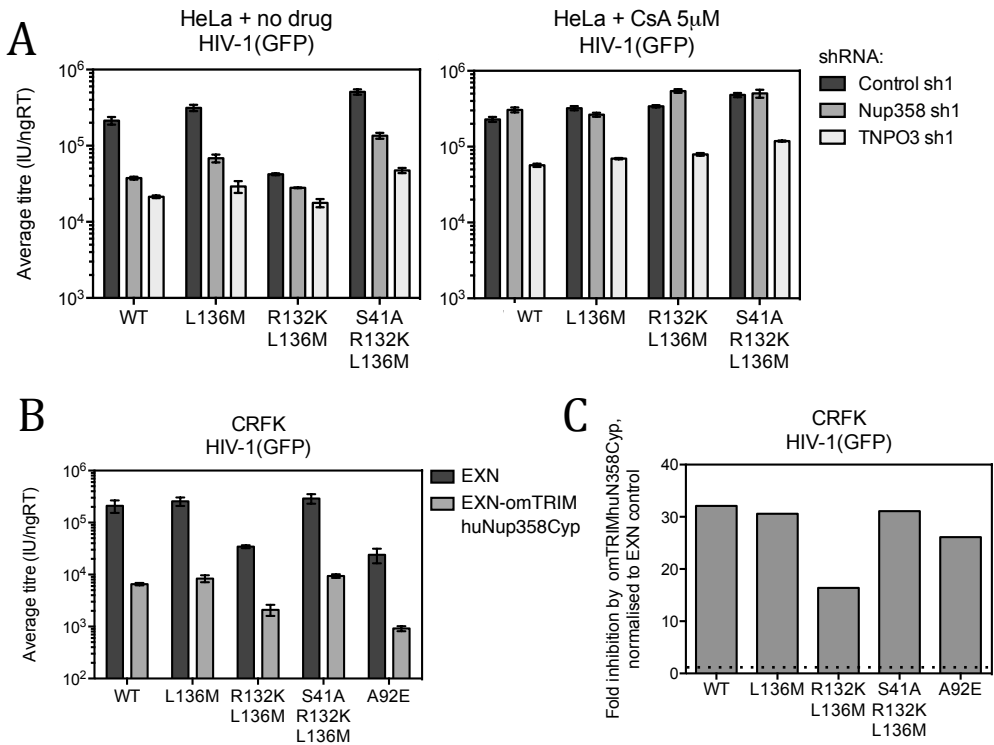
Figure 39. CsA restores sensitivity of R132K/L136M to TNPO3 depletion and R132K, but not A92E, mutation reduces the sensitivity of HIV-1 to omTRIMhuNup358Cyp

(A) HeLa cell clones stably expressing Nup358- or TNPO3-targeting shRNA or non-targeting control shRNA were infected with a titration of WT, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP) with (right panel) or without (left panel) CsA and viral titre measured 48hrs p.i. by flow cytometry. **(B)** CRFK cells stably expressing omTRIMhuN358Cyp or empty vector were infected with a titration of WT, A92E, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP) and viral titre was measured 48hrs p.i. by flow cytometry. **(C)** shows the data in (B) as fold difference between titre in control cells and titre in omTRIMhuN358Cyp-expressing cells. The dotted line shows where fold change = 1. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

4.2.6 R132K renders HIV-1 less sensitive to inhibition by HA-hCPSF6[68]-358, PF74, and BI-1, and S41A restores sensitivity to these antiviral drugs

In thus study we have demonstrated that CPSF6 dictates the nuclear entry pathway of HIV-1 (figure 23) and that R132K/L136M cannot use this nuclear entry pathway efficiently (figure 38A). We therefore sought to test whether this mutant differs in its ability to interact with CPSF6. As a surrogate assay for CPSF6 binding, we measured restriction by a CPSF6 C-terminal truncation mutant. We found that HA-hCPSF6[68]-358 restricts WT HIV-1 46-fold and L136M 57-fold, but that it only restricts R132K/L136M 10-fold (raw/fold data in figure 40A/B). Based upon these observations, one might hypothesise that R132K mutation causes the virus to infect independently of Nup358/TNPO3/Nup153 because it has a low affinity for CPSF6. It would follow that the compensatory mutation S41A, which restores the sensitivity of R132K/L136M to depletion of Nup358/TNPO3/Nup153, would also restore the affinity of CA for CPSF6. Unexpectedly, S41A/R132K/L136M was restricted by HA-hCPSF6[68]-358 to a similar degree as R132K/L136M, 13-fold. This suggests either that the independence of R132K/L136M upon Nup358/TNPO3 is not due to its reduced affinity for CPSF6 or, alternatively, that S41A/R132K/L136M is dependent upon Nup358/TNPO3 despite a low affinity for CPSF6. Either way, these observations suggest that dependence upon nuclear entry cofactors is not solely due to CPSF6. A discussed in section 1.5.2.5, CA-CypA interaction can also determine the dependence of HIV-1 upon Nup358/TNPO3 and so the reduced interaction between R132K/L136M CA and CypA is likely a contributing factor to this phenotype. In contrast to R132K/L136M, we found that A92E mutation, which we have shown does not affect dependence upon Nup358/TNPO3 (figure 38C), was restricted by HA-hCPSF6[68]-358 to a similar degree as WT HIV-1 (raw/fold data in figure 40C/D).

PF74 and BI-1 are antiviral compounds that bind to HIV-1 CA in the same conserved pocket as CPSF6 and interact with many of the same residues as CPSF6 (111, 327, 340) (these compounds are discussed further in section 5.1).
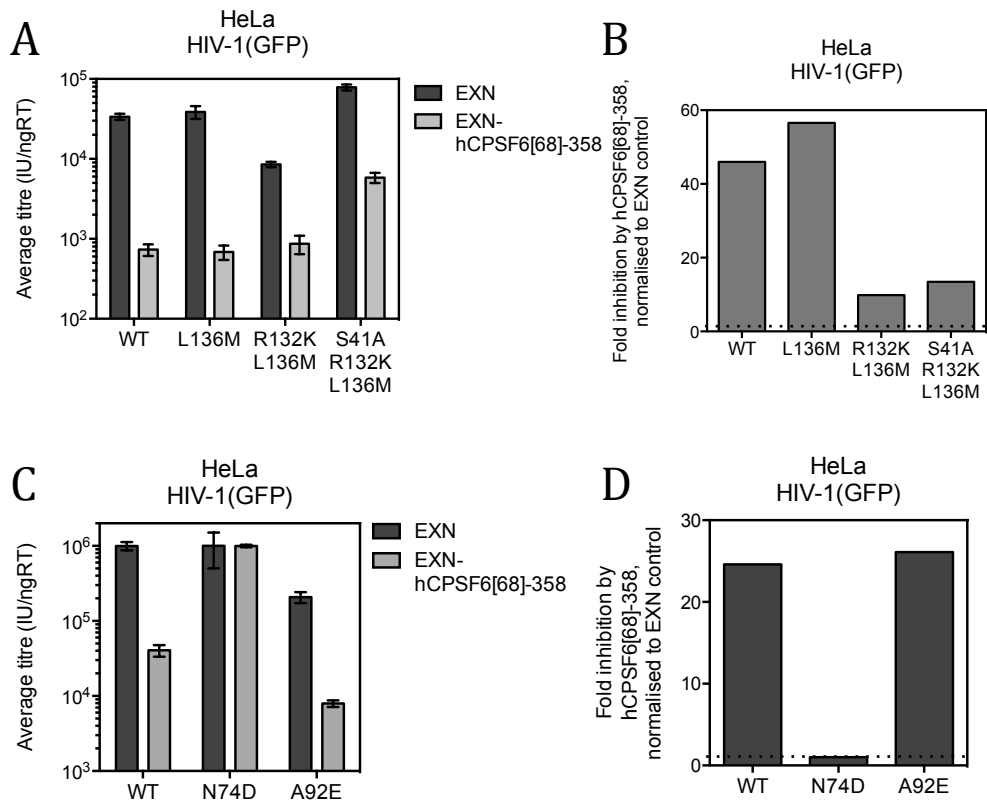
Figure 40. R132K, but not A92E, mutation in CA causes HIV-1 to become less sensitive to hCPSF6[68]-358 and S41A co-mutation does not restore sensitivity

HeLa cells stably expressing hCPSF6[68]-358 or empty vector were infected with a titration of **(A)** WT, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP) or **(C)** WT, N74D, or A92E HIV-1(GFP) and viral titre was measured 48hrs p.i. by flow cytometry. **(B)** and **(D)** show the data in (A) and (C), respectively, as fold difference between titre in control cells and titre in hCPSF6[68]-358-expressing cells. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

Like HA-hCPSF6[68]-358, inhibition by these CPSF6 peptidomimetics can be used as a way to measure changes in the structure of the CPSF6 binding pocket and so we sought to test the susceptibility of our B*27-associated mutants to inhibition by a titration of these compounds. We found that R132K/L136M was significantly less sensitive to inhibition by PF74 than WT or L136M (raw/normalised data in figure 41A/B). Similarly, R132K/L136M was completely insensitive to inhibition by BI-1 whilst WT and L136M were sensitive (raw/normalised data in figure 41C/D). Notably, whereas S41A did not restore sensitivity to HA-hCPSF6[68]-358, this compensatory mutation did restore sensitivity to both of these antiviral compounds. For example, 6.7μM PF74 inhibited WT, L136M, and S41A/R132K/L136M HIV-1 by 16 to 18-fold but only inhibited R132K/L136M 4-fold. Similarly, 90μM BI-1 inhibited WT, L136M, and S41A/R132K/L136M HIV-1 at least 7-fold but did not cause significant inhibition of R132K/L136M. Therefore, although S41A mutation did not restore the sensitivity of R132K/L136M to HA-hCPSF6[68]-358, it did restore the sensitivity of R132K/L136M to PF74/BI-1. These data suggest that both R132K and S41A mutations can significantly alter the structure of the CPSF6/PF74/BI-1-binding pocket of CA, as expected.

4.2.7 N74D mutation rescues R132K infectivity, but CPSF6 depletion does not

Having found that R132K mutation reduces sensitivity to inhibition by HA-hCPSF6[68]-358, PF74, or BI-1, but that the compensatory mutation S41A only restores sensitivity to inhibition by PF74 or BI-1, we probed the relationship between these mutations and CPSF6 further. It has previously been reported that the CA mutation A105T rescues the tire of R132K/L136M, like S41A does (326). Recently, A105T was identified as a CPSF6 binding mutant (247). We therefore tested whether the prototypical CPSF6 binding mutant, N74D, affected the titre of R132K/L136M. Indeed, we found that N74D rescued the infectivity of R132K/L136M (figure 42A). Based upon these observations, we hypothesised that the CA-CPSF6 interaction somehow inhibits R132K/L136M.
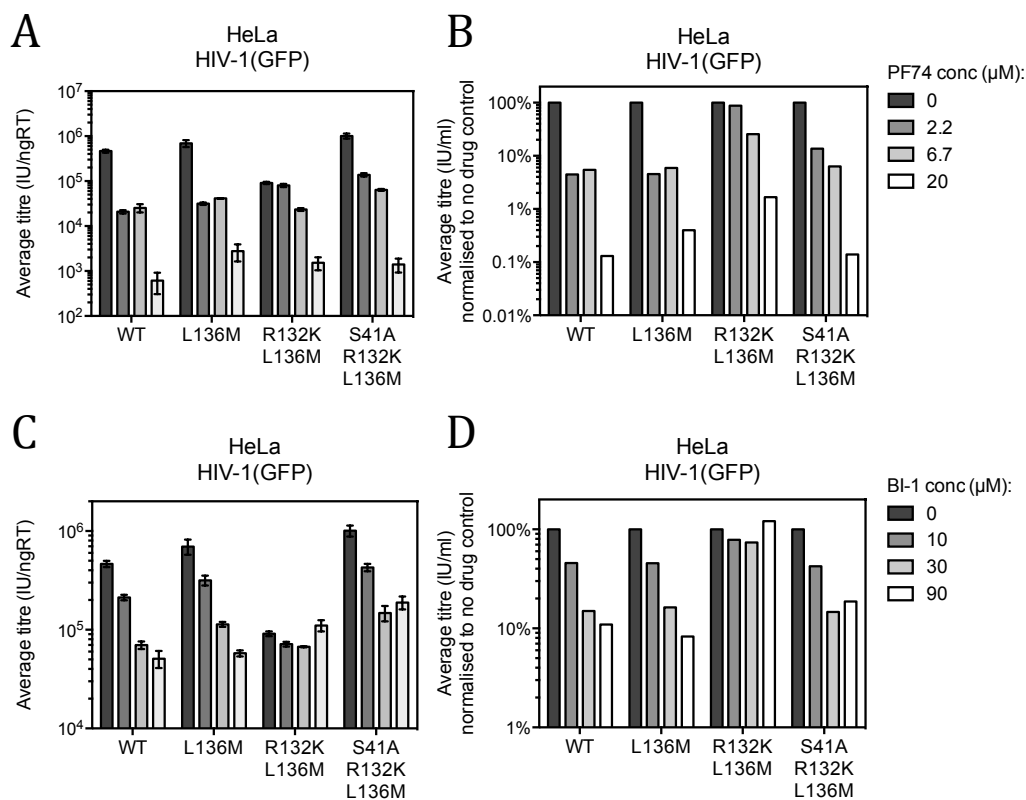
Figure 41. R132K mutation in CA causes HIV-1 to become less sensitive to PF74 and insensitive to BI-1 but S41A co-mutation restores sensitivity

HeLa cells were infected with a titration of WT, L136M, R132K/L136M, or S41A/ R132K/L136M HIV-1(GFP) in the presence of a titration of **(A)** PF74 or **(C)** BI-1 and viral titre was measured 48hrs p.i. by flow cytometry. **(B)** and **(D)** show the data in (A) and (C), respectively, normalised to viral titre in untreated cells. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).
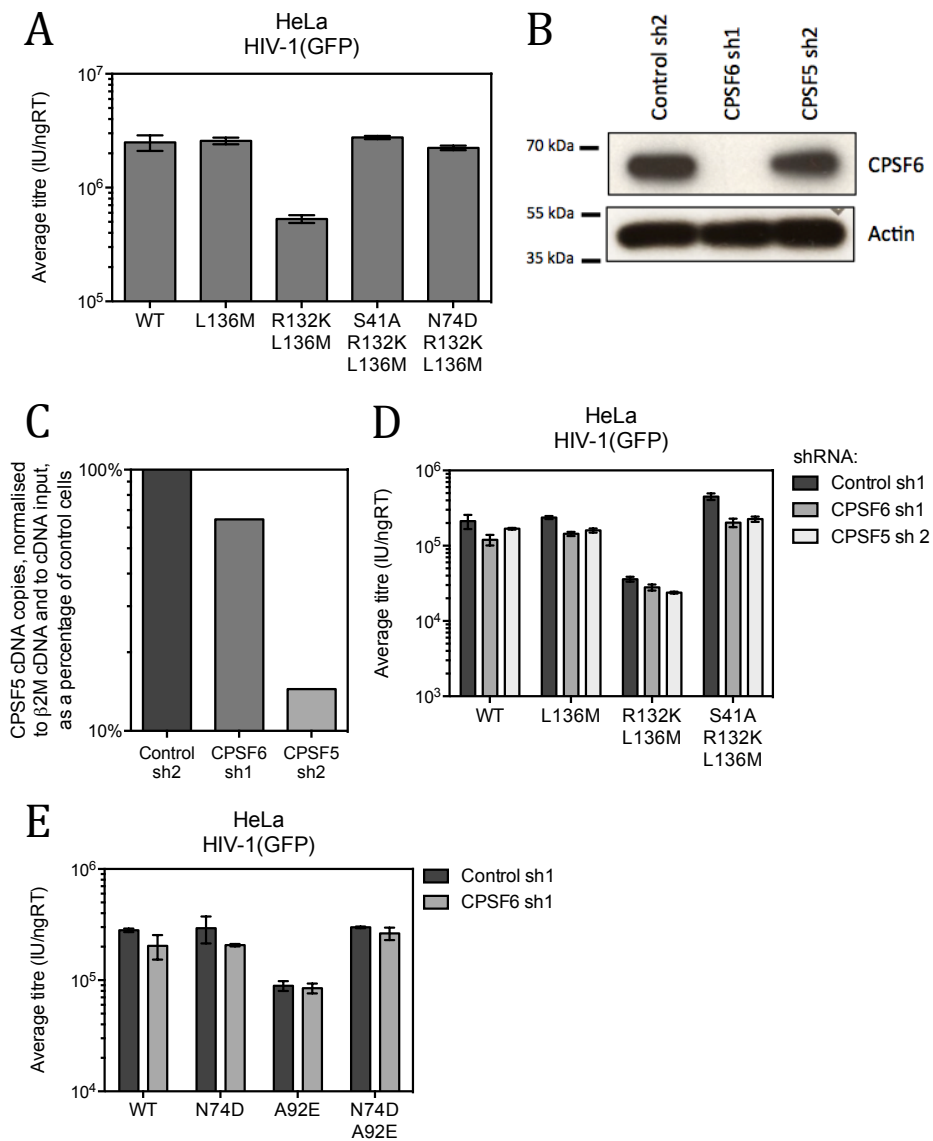
Figure 42. N74D rescues the infectivity defect caused by R132K but CPSF6 depletion does not

(A) Unmodified HeLa cells were infected with a titration of WT, L136M, R132K/ L136M, S41A/R132K/L136M, or N74D/R132K/L136M HIV-1(GFP) and viral titre measured at 48hrs p.i. by flow cytometry. **(B)** A western blot of lysates from HeLa cells transiently expressing CPSF6- or CPSF5-targeting shRNA or non-targeting control shRNA. β-actin serves as a loading control. **(C)** The extent of CPSF5 depletion in cells from (B) was measured by SYBR Green RT-qPCR of CPSF5 mRNA normalised to β2-microglobulin mRNA. **(D)** Cells in (B) and (C) were infected with a titration of WT, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP) and viral infectivity measured 48hrs p.i. by flow cytometry. **(E)** Cells in (B) were infected with a titration of WT, N74D, A92E, or N74D/A92E HIV-1(GFP) and viral infectivity measured 48hrs p.i. by flow cytometry. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

If R132K/L136M is inhibited for nuclear entry because it is unable to interact with certain cofactors of the normal nuclear entry pathway, a lack of interaction with CPSF6 may rescue the virus by releasing it from its commitment to this normal nuclear entry pathway. However, we found that neither efficient depletion of CPSF6 (figure 42B) nor of its binding partner CPSF5 (figure 42C) significantly affected the titre of R132K/L136M (figure 42D). Interestingly, we found that the titre of A92E, which has WT sensitivity to depletion of nuclear entry cofactors, is also rescued by co-mutation of N74D (figure 42E). This suggests that N74D mutation rescues the titre of these defective viruses for reasons that are unrelated to CPSF6 and Nup358/TNPO3 dependence.

4.2.8 HLA-B*27-associated mutations alter HIV-1 integration site selection

We have shown that the CA mutations R132K and S41A alter the way in which HIV-1 interacts with several components of the nuclear entry – Nup358, TNPO3, CypA, and CPSF6 – all of which have previously been shown to alter the integration site targeting of the virus (79, 159). We therefore characterised the integration sites of these HLA-B*27 associated mutants using the SureSelect[XT2] target enrichment system as described previously in section 3.2.7. Figure 43A shows the average values of the data used to created the dendrogram in figure 43B. L136M integrated into regions of 16.4 genes/Mbp and so clustered with WT HIV-1, which integrated into regions of 18.2 genes/Mbp. Consistent with our previous observations, R132K/L136M integrated into regions of 11.1 genes/Mbp, a significantly lower gene density than WT or L136M but still gene-rich compared to an average site in the human genome. In this chapter, we have shown that S41A/R132K/L136M has a WT phenotype with respect to infectivity as well as sensitivity to CsA, ApC, PF74, BI-1, omTRIMhuCypA, omTRIMhuN358Cyp, and depletion of Nup358 or TNPO3. Therefore, one may predict that it would also phenocopy WT HIV-1 with respect to integration site targeting. Unexpectedly, it integrates into regions with only 8.5 genes/Mbp and so it clusters with N74D, which integrates at a gene density equivalent to

A

Name	% in gene	Genes/Mb ⬇	CpG islands/Mb
HIV-1	78.5	18.2	25.8
HIV-1 L136M	84.4	16.4	22.4
HIV-1 R132K, L136M	73.4	11.1	14.5
HIV-1 S41A, R132K, L136M	71.6	8.5	8.6
N74D	70.4	7.0	6.1

B
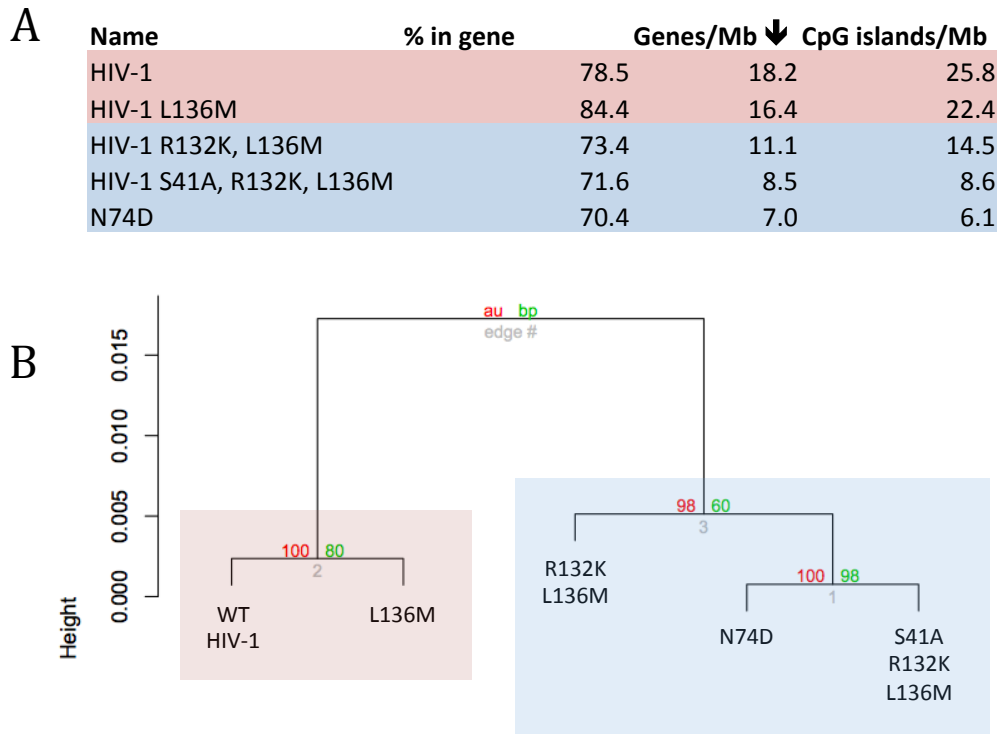
Figure 43. B*27-associated mutations alter the integration site selection of HIV-1

Unmodified HeLa cells were infected with WT, N74D, L136M, R132K/L136M, or S41A/R132K/L136M HIV-1(GFP) at an MOI of ~1.0. At 24hrs p.i. DNA was extracted, sheared, enriched for fragments containing HIV-1 LTRs using the SureSelect[XT2], and sequenced. **(A)** The results of integration site analysis for gene density, CpG-island density, and the proportion of proviruses within genes, sorted by gene density as indicated by the black arrow. **(B)** A dendrogram based upon the gene density and CpG island density values shown in (A). The y-axis of the dendrogram, Height, uses arbitrary units to represent the relative closeness of either individual data points or of clusters. The nodes are statistically supported by bootstrap probability (bp) *p* values shown in green and approximately unbiased (AU) *p* values shown in red. The red/blue shading in C and D highlights the two major clusters formed by these samples.

random. The only way in which we have found S41A/R132K/L136M to not phenocopy WT HIV-1 is with respect to HA-hCPSF6[68]-358-mediated restriction. This could be interpreted to mean that the inability of this mutant to integrate into regions enriched in genes is due to a lack of interaction with CPSF6. Instead, we hypothesise that this observation is misleading and that S41A restores the affinity of CA for CPSF6, as it appears to do for PF74 and BI-1. We therefore propose that there are other factors involved in integration site targeting of HIV-1 that we have not yet identified.

4.2.9 The CA context of HLA-B*27-associated mutations can alter their phenotype

Thus far, we have carried out all experiments using a single viral strain and so we sought to test whether the phenotypes of these HLA-B*27 associated mutations are context dependent. We used four viruses (provided by Yamashita. M) that were created by cloning CA from viral isolates from four different HLA-B*27[+] patients into a pBru3oriΔEnvGFP3 backbone, named CR0339X, PRLS24, CR0206U, and CR0312W (335). As depicted in the CA alignment in figure 44A, all four viruses encode M/I at position 136 and K at position 132, the typical HLA-B*27 escape mutations in the epitope KK10. Due to the extremely poor replication of viruses containing the R132K mutation in the absence of any compensatory mutation, one would expect that the vast majority of clinical isolates containing R132K would also be mutated at residue 41. As expected, these viruses had mutated residue 41 from S to either A (CR0339X and PRLS24), V (CR0206U), or T (CR0312W). We sought to compare the phenotypes of these viruses to the virus used in this study that contains CTL escape and compensatory mutations, S41A/R132K/L136M.

For each of the 5 viruses described above, we used SDM to revert residue 41 to the wild-type non-compensatory serine and tested the effect of this upon viral infectivity. As expected, reversion of residue 41 to an S caused a significant decrease in titre for all viruses. However, the magnitude of this effect was context dependent and thus ranged from 4 to 21-fold (figure 44B). CR0312W

A

41 ↓

```
CR0339X                    PIVQNAQGQMVHQAISPRTLNAWVKVVEEKAFNPEVIPMFAALSEGATPQ  50
CR0206U                    PIVQNMQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFVALSEGATPQ  50
PRLS24                     PIVQNLQGQMVHQPLSPRTLNAWVKVIEEKAFNPEVIPMFAALSEGATPQ  50
CR0312W                    PIIQNLQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFTALSEGATPH  50
p8.91Ex S41A:R132K:L136M   PIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFAALSEGATPQ  50
Score                       **  **  *******  *********** ****  ******  ********
```

```
CR0339X                    DLNTMLNTVGGHQAAMQILKETINEEAAEWDRLHPVHAGPIAPGQMREPR  100
CR0206U                    DLNTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGPNPPGQMREPR  100
PRLS24                     DLNTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGPIAPGQMREPR  100
CR0312W                    DLNTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGPIAPGQMREPR  100
p8.91Ex S41A:R132K:L136M   DLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPR  100
Score                       *****************  ***************  ******    ********
```

132 ↓   **136** ↓

```
CR0339X                    GSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKKWIIMGLNKIVKMYSPISI  150
CR0206U                    GSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKKWIIGLNKIVRMYSPTSI  150
PRLS24                     GSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKKWIIMGLNKIVRMYSPVSI  150
CR0312W                    GSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKKWIIMGLNKIVRMYSPVSI  150
p8.91Ex S41A:R132K:L136M   GSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKKWIIMGLNKIVRMYSPTSI  150
Score                       **************** ***************** ****** **** **
```

```
CR0339X                    LDIROGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKT  200
CR0206U                    LDIROGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKT  200
PRLS24                     LDIROGPKEPFRDYVDRFYKTLRAEQATQEVKNWMTETLLVQNANPDCKT  200
CR0312W                    LDIROGPKEPFREYVDRFYKTLRAEQASQEVKNWMTETLLIQNANPDCKT  200
p8.91Ex S41A:R132K:L136M   LDIROGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKT  200
Score                       ***********  *************** *  **********  *********
```

```
CR0339X                    ILKALGPAATLEEMMTACQGVGGPGHKARVLA  232
CR0206U                    ILKALGPAATLEEMMTACQGVGGPSHKARVLA  232
PRLS24                     ILKALGPAATLEEMMTACQGVGGPSHKARVLA  232
CR0312W                    ILKALGPGATLEEMMSACQGVGGPGHKARVLA  232
p8.91Ex S41A:R132K:L136M   ILKALGPGATLEEMMTACQGVGGPGHKARVLA  232
Score                       ******* ******* ******** ******
```

B

HeLa
HIV-1(pBruOriΔEnv-GFP)

Figure 44. The CA context of B*27-associated mutations influences their effect on viral titre
(A) An amino acid alignment of CA encoded by pBru3oriΔEnvGFP3-based CR0339X, PRLS24, CR0206U, or CR0312W (where CA was derived from clinical isolates) or the p8.91Ex-based S41A/R132K/L136M HIV-1(GFP). B*27-associated residues are highlighted with black arrows. Asterisks highlight residues that are conserved in all 5 sequences. **(B)** Unmodified HeLa cells were infected with a titration of S41A/R132K/L136M, R132K/L136M, CR0339X, CR0339X A41S, PRLS24, PRLS24 A41S, CR0206U, CR0206U V41S, CR0312W, or CR0312W T41S and viral titre measured 48hrs p.i. by flow cytometry. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

198

encodes threonine at position 41, which is found in 25% of all CA sequences in the Los Alamos database (http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html), has a polar uncharged side chain like serine, and has not been shown to compensate for R132K mutation. The finding that reversion of this threonine to serine exhibited a very small decrease in titre could mean that S41T is a poor compensatory mutation and/or that another unknown compensatory mutation is present in this virus's CA.

We next tested whether the CsA-dependence of L136M/R132K was context dependent. Figure 45A shows that whilst all viruses encoding A/V/T41 were insensitive to 5μM CsA treatment, CsA rescued the reduced titres of all the viruses encoding S41. The degree of rescue conferred by CsA was context dependent. As before, CR0312W had an especially weak phenotype; CsA only rescued the titre of CR0312W by 2-fold, whilst it rescued all others 7 to 14-fold. We also used this panel of viruses to test whether the ApC-sensitivity of R132K/L136M was context dependent. Figure 45B shows that the titres of all viruses encoding A/V/T41 were unchanged by ApC-mediated cell cycle arrest. In contrast, 2μg/ml ApC significantly inhibited all viruses encoding S41, suggesting they cannot efficiently enter the nucleus via the NPC and are dependent upon mitosis. The degree of inhibition was context dependent. Again, CR0312W exhibited an exceptionally weak phenotype and was only inhibited 4-fold, whilst all other viruses were inhibited between 21 and 130-fold. These observations suggest that a compensatory mutation other than S41T exists in CR0312W.

Having previously found that R132K/L136M was less sensitive to Nup358/TNPO3 depletion than WT HIV-1 but that the compensatory mutation S41A restored sensitivity, we sought to test whether this phenotype was context dependent. Figure 46A shows that all viruses were significantly more sensitive to depletion of these two nuclear entry cofactors when encoding A/V41 than when encoding S41, as expected. In contrast, CR0312W was sensitive to the depletion of these proteins regardless of whether it encoded T or S at residue 41. The finding that CR0312W (encoding T41) does not have
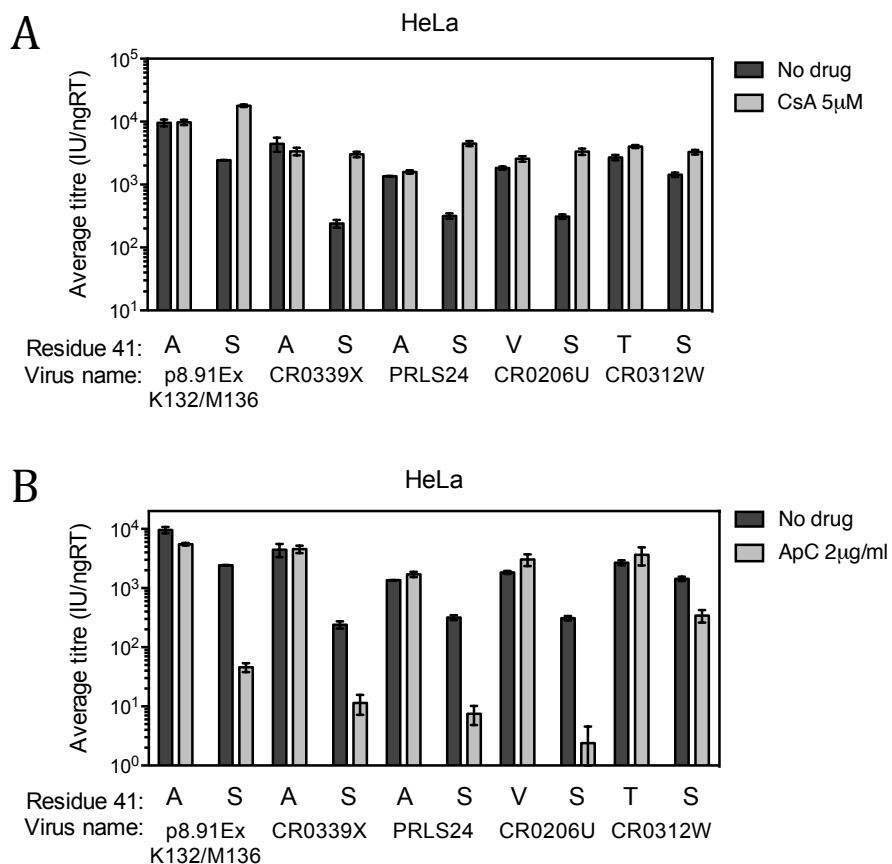
Figure 45. The CA context of B*27-associated mutations determines their relationship with CsA and ApC

HeLa cells were infected with a titration of S41A/R132K/L136M, R132K/L136M, CR0339X, CR0339X A41S, PRLS24, PRLS24 A41S, CR0206U, CR0206U V41S, CR0312W, or CR0312W T41S **(A)** +/- 5μM CsA or **(B)** +/- 2μg/ml ApC and viral titre measured 48hrs p.i. by flow cytometry. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).
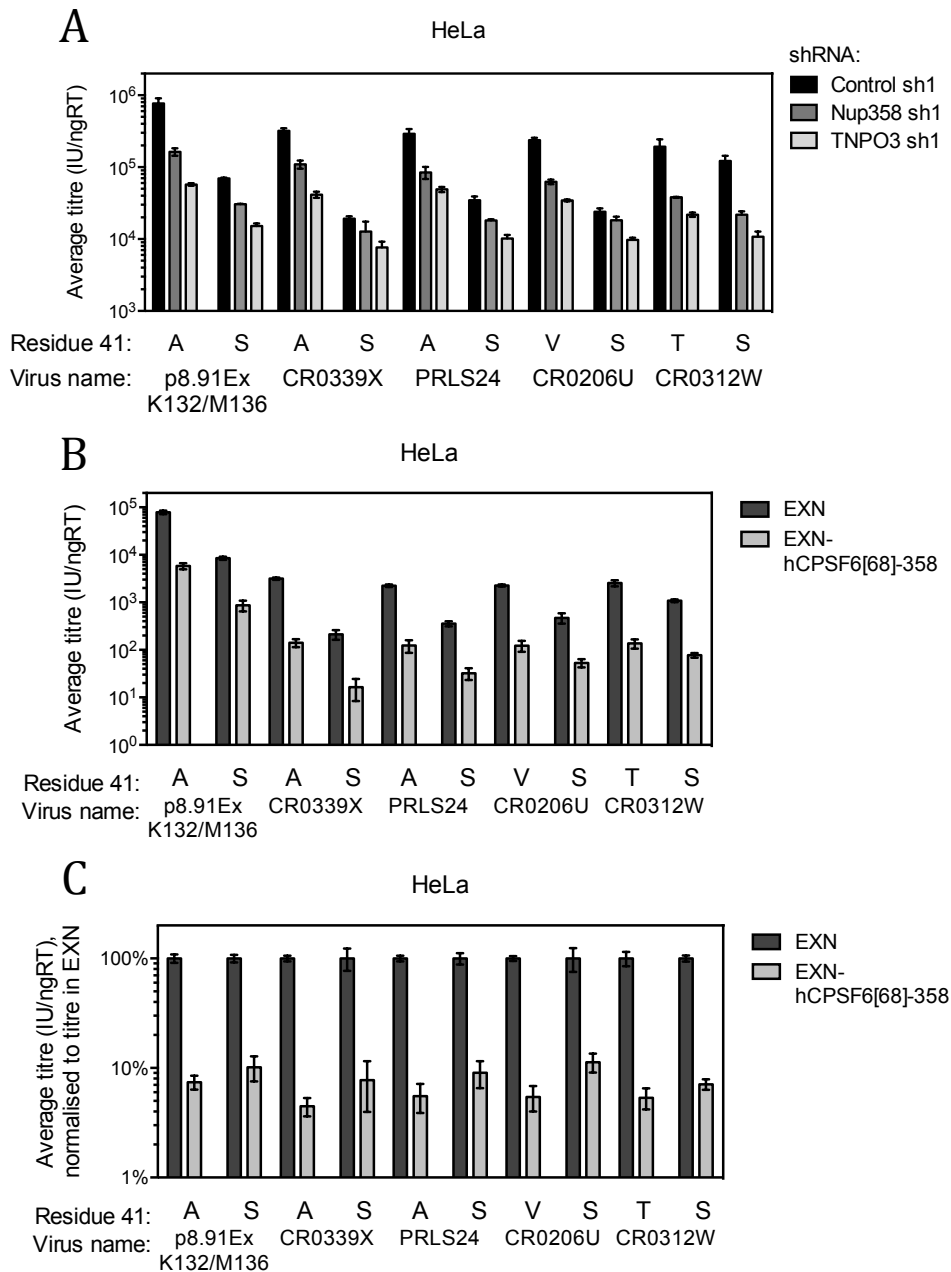
Figure 46. The CA context of B*27-associated mutations determines their sensitivity to Nup358/TNPO3 depletion and to hCPSF6[68]-358

(A) HeLa cell clones stably expressing Nup358- or TNPO3-targeting shRNA or non-targeting control shRNA or **(B)** HeLa cells stably expressing hCPSF6[68]-358 or empty vector were infected with a titration of S41A/R132K/L136M, R132K/L136M, CR0339X, CR0339X A41S, PRLS24, PRLS24 A41S, CR0206U, CR0206U V41S, CR0312W, or CR0312W T41S and viral titre measured 48hrs p.i. by flow cytometry. **(C)** Data in (B) normalised to titre in EXN cells. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

reduced dependence upon these cofactors suggests that at least one other residue in CA is acting as a compensatory mutation for the CTL escape mutation R132K.

We next measured whether the sensitivity of these viruses to Nup358/TNPO3 depletion correlated with their sensitivity to HA-hCPSF6[68]-358-mediated restriction, which is a surrogate assay for the affinity of CA for CPSF6. Figure 46B/C (raw/normalised data, respectively) shows that the residue encoded at position 41 did not significantly alter the level of inhibition conferred by this artificial restriction factor; all viruses were restricted between 10 and 22-fold. This implies that the affinity of HIV-1 CA for CPSF6 is not altered by S41A/V/T mutation. Therefore, the affinity of CA for CPSF6 cannot explain the restoration of Nup358/TNPO3 dependence conferred by S41A/V/T mutation.

4.3 Discussion

CTL escape mutations associated with the protective HLA alleles B*57 and B*27 have been reported to alter the relationship of the virus with CypA (325, 326). In light of the finding that the CA-CypA interaction can dictate the nuclear entry pathway of HIV-1 (79), we hypothesised that these CTL escape mutations may differ in their nuclear entry pathway. This would be of great interest as it would inform us about the significance of this nuclear entry pathway to viruses containing clinically relevant mutations. Whilst we found that the B*57-associated mutations did not significantly affect the nuclear entry pathway of HIV-1, we found that the B*27-associated CTL escape mutation in CA R132K is defective in its use of the CPSF6-dependent nuclear entry pathway described in section 3.3.8. Intriguingly, we found that the compensatory mutation S41A restores the ability of the virus to interact with all components of the nuclear entry that we measured except for HA-hCPSF6[68]-358, as is discussed below.

4.3.1 B*57-associated CA mutations do not alter the nuclear entry pathway of HIV-1

Firstly we examined CA mutations in the epitope TW10, which are associated with escape from the protective HLA allele B*57. In agreement with Brockman, et al. (325), we found that the proposed CTL escape mutations T110N and G116A were more sensitive to CsA-mediated inhibition than WT HIV-1. In contrast, the proposed compensatory mutations H87Q, I91L, and M96I did not alter the relationship of the virus with CsA, despite their location within or close to the CypA binding loop of CA. These findings lead us to hypothesise that other components of the nuclear entry pathway may also be affected by these mutations. We found that T110N, I91L, and M96I were less sensitive than WT HIV-1 to depletion of Nup358. However, given that these CA mutants retained sensitivity to depletion of TNPO3 and to HA-hCPSF6[68]-358, we propose that they still enter the nucleus via the same pathway at WT HIV-1, albeit less efficiently. These mutations lie within or proximal to the CypA-binding loop of CA, which interacts directly with $Nup358_{Cyp}$ (79). We therefore hypothesised that their insensitivity to Nup358 depletion is due to a reduced ability to bind to

this nucleoporin. None of these B*57-associated mutants were inhibited by ApC-mediated cell-cycle arrest, indicating that they were all able to infect non-dividing cells via the NPC. We therefore interpret these findings to mean that, although some of these B*57-associated mutations may slightly disrupt interaction with CypA and/or Nup358, they do not dramatically affect the nuclear entry pathway of the virus.

4.3.2 B*27-associated CA mutations alter the nuclear entry pathway of HIV-1

In contrast to the B*57-associated mutations in the epitope TW10, we found that some HLA-B*27-associated mutations in KK10 do significantly alter the nuclear entry pathway of HIV-1. L136M did not significantly alter any of the nuclear entry phenotypes measured. In contrast to L136M, and in agreement with Schneidewind, et al. (326), we found that R132K causes a significant reduction in viral titre. We found the magnitude of this infectivity defect to be cell-type dependent, ranging from 3-fold in TE671s to 252-fold in CEM cells. Importantly, we also observed a 5-fold defect in single-round infection of primary CD4$^+$ T cells, confirming the importance of this mutation in the cell type most commonly infected by HIV-1 *in vivo*. The virus evolves R132K due to a strong selective pressure to escape the CTL response, but the virus is then under pressure to replicate more efficiently and so it evolves the compensatory mutation S41A. Like Schneidewind *et al.* we found that S41A rescued the infectivity defect caused by R132K in all cell types, no matter how large the defect was, such that S41/R132K/L136M was as infectious as WT HIV-1 (326). The ability of this triple mutant to escape the CTL response and to replicate efficiently explains why its detection *in vivo* usually coincides with disease progression (381).

We found R132K/L136M infection of HeLa cells to be defective after reverse transcription but prior to 2-LTR-circle formation, which suggests that the virus is defective in nuclear entry. A previous report showed that R132K infection was defective after nuclear entry (403) but the qPCR probes used to measure 2-

LTR circles in this study (and in many other studies) also measure autointegrants, which can form in the cytoplasm and do therefore not signify nuclear entry (247) (as discussed in section 1.5.1.1). Therefore, in this study, we used probes that specifically recognise the LTR-LTR junction of 2-LTR circles and in doing so revealed that R132K/L136M is defective prior to/at the stage of nuclear entry.

The infectivity defect of R132K was reportedly rescued by CsA (326), a phenotype we recapitulated in this study. We found the infectivity of WT, L136M, R132K/L136M, and S41A/R132K/L136M to be equivalent in the presence of 5μM CsA. In cell types such as HeLa and CEM, in which WT HIV-1 is unaffected by CsA, this meant that the titre of R132K/L136M was increased and this mutant is therefore considered CsA-dependent. In cell types such as Jurkat and TE671, in which WT HIV-1 is inhibited by CsA, R132K/L136M was unaffected by CsA and is therefore considered CsA-resistant. This CsA-resistant/dependent phenotype has previously been reported for the CA mutation A92E, which was selected for by serial passage of HIV-1 in the presence of CsA (297) but has never been found to arise *in vivo*. In this study we confirmed that A92E is indeed CsA-dependent in HeLa cells, like R132K/L136M. As discussed in section 1.5.2.3, the infectivity defect of A92E in TE671s can be rescued by overexpression of CypA, suggesting that the defect may be simply due to CypA concentration (298). However, neither a gene expression analysis of six different human cell lines (HeLa, H9, CEM, HOS, Jurkat, and 293T) (300) nor a screen of 32 different HeLa cell clones (112) found a correlation between permissivity to A92E and expression of CypA. This CsA-resistant/dependent phenotype has therefore been attributed to an unidentified CypA-dependent restriction factor (299). This putative restriction factor may be difficult to identify if its potency depends not only on its expression level but also the expression levels of CypA. Further investigation is therefore required to determine whether this restriction factor exists, whether it is responsible for the infectivity defects of A92E and/or R132K/L136M and, if so, how S41A escapes this restriction.

Given the unusual relationship of R132K/L136M and A92E with CypA/CsA, we asked whether either of these mutations alters the affinity of CA for CypA. Despite residue A92 lying within the CypA binding loop of CA, A92E CA has previously been reported to interact with CypA with an affinity equivalent to WT CA (298). In contrast to A92, both R132 and S41 lie within the CPSF6-binding pocket, in helix 7 and 2 respectively. Despite their distance from the CypA binding loop, mutations in the CPSF6-binding pocket have previously been reported to alter interaction with CypA *in vitro* (358). As a surrogate measure of CypA binding, we measured the effect of these mutations upon the level of restriction conferred by omTRIMhuCypA, which binds to cores via its CypA domain. In agreement with previous reports, A92E mutation did not alter the level of restriction but, intriguingly, R132K mutation reduced the level of restriction from 47-fold to 25-fold and subsequent S41A mutation restored it to 44-fold. Importantly, this 2-fold reduction in omTRIMhuCypA-mediated restriction does not necessarily represent a 2-fold decrease in the affinity of R132K CA for CypA. TRIM5-based restriction factors such as this exist as dimers and multimerisation of CypA has been shown to increase its avidity for CA (405). Therefore, the 2-fold reduction in restriction caused by R132K mutation may actually represent a much more significant reduction in the affinity of CA for CypA.

As well as being CsA-resistant/dependent, A92E is reportedly sensitive to ApC-mediated cell-cycle arrest (298). In agreement with this we found that ApC inhibits A92E, but not WT HIV-1, in HeLa cells. Similarly, we found that ApC inhibits R132K/L136M but not L136M or S41A/R132K/L136M in HeLa cells. In fact, R132K/L136M was significantly more sensitive to ApC than the other viruses in all cell lines tested, although the degree of inhibition varied. These observations suggest that A92E or R132K mutation prevents HIV-1 from being able to enter the nucleus via the NPC and instead must enter the nucleus during mitosis. Importantly, we have shown that CsA rescues A92E and R132K from inhibition by ApC, suggesting that it is the CA-CypA interaction that prevents these mutants from traversing the NPC efficiently. The fact that we do not know why A92E or R132K are defective/CsA-dependent makes this observation

difficult to interpret. However, whether these mutants are inhibited by CypA or by a CypA-dependent restriction factor, we propose that the inhibition is specific to NPC-mediated nuclear import. Unexpectedly, it was found that A92E (298) and R132K/L136M (Rasaiyaah. J, unpublished observations) replicate as efficiently as WT HIV-1 in non-dividing primary MDMs. This suggests that HeLa cells and MDMs differ in their expression levels of CypA and/or the putative restriction factor.

The finding that A92E or R132K mutation alters the relationship of HIV-1 with CypA lead us to hypothesise that these mutations may also alter the relationship of the virus with downstream nuclear entry cofactors whose use is dependent upon CA-CypA interaction, namely Nup358, TNPO3, and Nup153 (79). Indeed, we found that R132K mutation significantly reduced the sensitivity of the virus to depletion of each of these cofactors and that co-mutation of S41A restored sensitivity. However, the reduced sensitivity of R132K/L136M to TNPO3 depletion cannot be due to its reduced interaction with CypA because CsA treatment (which would reduce the CA-CypA interaction even further) caused R132K/L136M to become more sensitive to TNPO3 depletion.

We therefore considered the possibility that these B*27-associated mutations could alter the sensitivity of HIV-1 to depletion of each of these cofactors separately. With respect to Nup358, we found that R132K mutation reduced sensitivity to omTRIMhuNup358Cyp and that the S41A restored sensitivity. With respect to Nup153, it has recently been reported that $Nup153_{1410-1418}$ interacts directly with CA residues L136 and S41 (when in complex with hexameric CA) and lies in close proximity to R132 (83, 340). It is therefore possible that the mutations R132K and S41A could reduce and enhance the ability of CA to interact with Nup153, respectively. In other words, R132K may reduce the sensitivity of HIV-1 to Nup358 and Nup153 depletion because it cannot interact with these nucleoporins anyway. However, we cannot explain the reduced sensitivity of R132K/L136M to TNPO3 depletion in the same way.

We therefore sought to test the hypothesis that R132K alters the sensitivity to HIV-1 to these nuclear entry cofactors by reducing the affinity of CA for CPSF6. As discussed in section 3.3, it has recently been shown that CPSF6[68]$_{276-290}$ not only interacts with helices 3 and 4 of one CA monomer but also with helices 2, 7, 8, and 9 of the adjacent CA monomer in a hexamer (340). This 15mer of CPSF6 does not interact directly with residues L136, R132, or S41, but it lies in close proximity to them. It is therefore plausible that mutation of these residues would alter CA-CPSF6 interaction, particularly given that full-length CPSF6 would make much more extensive contacts with CA than this 15mer. As surrogate assays of CPSF6 binding, we measured inhibition of the B*27-associated mutants by HA-hCPSF6[68]-358 as well as two small molecule inhibitors of HIV-1, PF74 and BI-1, which interact with the same CA pocket as CPSF6 (111, 327, 340) (discussed further in section 5.1). As hypothesised, we found that R132K mutation significantly reduced the sensitivity of HIV-1 to HA-hCPSF6[68]-358 and PF74 and made it completely insensitive to BI-1. Furthermore, S41A mutation restored sensitivity to PF74 and BI-1. Unexpectedly, S41A did not restore sensitivity to HA-hCPSF6[68]-358. One might expect that, of HA-hCPSF6[68]-358, PF74, and BI-1, the interaction of CA with HA-hCPSF6[68]-358 would most closely mimic its interaction with endogenous CPSF6. However, given that the structure of full-length CPSF6 remains unknown, there is no evidence to support this and it is possible that such a large truncation of CPSF6 significantly alters the structure of the protein. Importantly, the observations made here clearly demonstrate that both R132K and S41A alter the structure of this pocket in CA. We therefore hypothesise that R132K mutation reduces the sensitivity of HIV-1 to depletion of Nup358, TNPO3, and Nup153 because it reduces the affinity of CA for CPSF6. Similarly, we propose that S41A restores sensitivity to depletion of these nuclear entry cofactors because it restores the affinity of CA for CPSF6 to WT levels. In contrast to R132K, we found that A92E mutation does not affect the sensitivity of HIV-1 to depletion of TNPO3 or Nup358, nor to omTRIMhuN358Cyp or HA-hCPSF6[68]-358. Therefore, despite its unusual relationship with CypA, A92E appears to interact normally with all other known components of the HIV-1 nuclear entry pathway. The fact that both A92E and R132K are CsA-dependent

in HeLa cells and yet only R132K has an altered relationship with CPSF6, Nup358, TNPO3, and Nup153 allows us to differentiate between the two phenotypes. In other words, it reveals that being CsA-dependent does not necessarily confer independence of the rest of the HIV-1 nuclear entry pathway.

Our findings thus far suggest that HIV-1 can enter the nucleus via two different nuclear entry pathways. One pathway is taken by WT HIV-1 and is dependent upon CPSF6, Nup358, TNPO3, and Nup153. The other is taken by N74D and is completely independent of these cofactors. Both of these pathways are equally efficient in HeLa cells. The findings that S41A rescues the infectivity defect of R132K/L136M and also restores its ability to efficiently use the CPSF6-dependent pathway lead us to the hypothesis that partial or inefficient use of the CPSF6-dependent pathway causes an infectivity defect. We therefore hypothesised that allowing R132K/L136M to use the CPSF6-independent pathway may also rescue its infectivity defect. Whilst we found that N74D rescued R132K/L136M infectivity, we were intrigued to find that CPSF6 depletion did not. Henning, et al. (335) recently reported that, in their hands, CPSF6 depletion does partially rescue the titre of R132K/L136M in HeLa cells. Whilst it is possible that they achieved a stronger depletion than ours, the level of depletion we achieved was sufficient to make WT HIV-1 insensitive to depletion of TNPO3 or Nup358. Another possible explanation for our observation is that N74D mutation confers an unknown phenotype that is unrelated to CPSF6 binding. This notion is supported by the finding that N74D similarly rescues the titre of A92E, which we have shown is able to use the CPSF6-dependent nuclear entry pathway efficiently.

In light of our findings that B*27-associated mutations alter the nuclear entry pathway of HIV-1, we investigated whether they also affect the integration site targeting of the virus. As discussed in section 1.5.2.5, disrupting CA-CypA interaction causes HIV-1 to integrate into regions of a higher gene density (79). In contrast, disrupting the CA-CPSF6 interaction or depleting Nup358 or TNPO3 leads to integration into regions of a lower gene density (figure 25)(79, 159). Furthermore, it was shown that CA-CPSF6 interaction was dominant over CA-

CypA interaction with respect to integration site selection (79). Consistent with these observations, we found that R132K/L136M integrates into regions of a significantly lower gene density than WT HIV-1 and L136M. However, S41A/R132K/L136M unexpectedly integrated into regions of an even lower gene density, approaching the phenotype of N74D. Based on this observation, we hypothesised that that there are other host cofactors involved in HIV-1 integration site selection that have not yet been identified.

As discussed in section 1.2.9.6, it has been widely proposed that integration into gene-rich regions is beneficial to HIV-1 because they are more transcriptionally active and so proviral transcription would be increased. Conversely, integration into regions of a lower gene density would result in lower proviral transcription, lower virus production, and slower replication. However, there are several observations that strongly argue against this hypothesis. For example, S41A/R132K/L136M has been shown to replicate efficiently in primary PBMCs, CD4$^+$ T cells (326), and MDMs (Rasaiyaah. J, unpublished observations). Furthermore, its appearance *in vivo* is associated with an increase in viral load and disease progression (326). A longitudinal study of three HLA B*27 negative patients who each contracted HIV-1 carrying the mutations S41A, R132K, and L136M from a B*27 positive patient revealed that these mutations did not revert inside their new HLA-mismatched hosts even after 1-3 years of active viral replication (patients were all treatment naïve) (406). In contrast, the B*57-associated TW10 mutations tested in this study were found to revert within a few months of transmission to an HLA-mismatched host (395). These observations suggest that integration into gene dense regions is not necessary for efficient viral replication *in vitro* or *in vivo*. However, it is possible that the virus developed further compensatory mutations that allowed it to integrate into gene rich regions *in vivo*. Importantly, none of the mutations tested here altered the ability of the virus to integrate inside genes, another characteristic that is thought to be necessary for efficient proviral transcription (detailed in section 1.2.9.3).

In order to further assess the consequences of these B*27-associated mutations arising *in vivo*, we examined whether the CA context of these mutations affected their phenotypes. We compared the viral strain used throughout this study (p8.91Ex S41A/R132K/L136M) with 4 viruses containing CA derived from B*27+ patients. These clinical CA sequences all encoded the mutations S41A/V/T, R132K, and L136M/I, differed from p8.91Ex by 9-15 residues, and from each other by 2-8 residues. Whilst S41A is a known compensatory mutation of R132K, S41T occurs in 25% of HIV-1 sequences and S41V has not previously been observed and so it is unknown whether they serve as compensatory mutations or not.

For each virus, we reverted residue 41 to serine and assessed its effect upon the nuclear entry pathway of HIV-1. For three of the four viruses – CR0339X, PRLS24, and CR0206U – reversion of residue 41 to serine resulted in very similar phenotypes as we had previously observed; they became significantly less infectious, dependent upon CsA, sensitive to ApC, and less sensitive to depletion of Nup358 and TNPO3. These findings inform us that, like S41A, S41V is an efficient compensatory mutation for R132K. Furthermore, the magnitude of the effect of these mutations is dependent on the CA context.

In contrast to these three viruses, CR0312W exhibited very small changes in phenotype, if any, when residue 41 was reverted from threonine to serine. It became only 3-fold less infectious, was rescued only 2-fold by CsA, was inhibited only 4-fold by ApC, and its sensitivity to depletion of Nup358 and TNPO3 was unchanged. In other words, CR0312W is phenotypically similar to p8.91Ex S41A/R132K/L136M regardless of the residue at position 41. We interpret this to mean that another unidentified residue within CR0312W is compensating for R132K. There are 7 residues within CR0312W CA that differ from all other isolates and could thus be compensating for R132K: I6, H50, E163, D180, I191, G208, and S216. Two residues of particular interest are residue 6, which lies in the β-hairpin loop of CA_{NTD} and may therefore influence the CypA-binding loop, and residue 50, which lies in helix 3 of the CPSF6 binding pocket close to S41. Further mutational studies would be needed to

determine whether any of these mutations, or indeed a combination of them, compensates for R132K. From these data, it is not possible to determine whether S41T mutation would be able to compensate for R132K (if the other unidentified mutation were not present). However, residue 41 is a threonine in 25% of isolates and, if this were able to compensate well for R132K mutation, then S41A would not be selected for as often as it is nor would the unidentified compensatory mutation. We therefore conclude that S41T is a very poor compensatory mutation for R132K.

As detailed in section 3.3.7, Rasaiyaah, et al. (92) recently demonstrated that both CypA and CPSF6 are required for HIV-1 to evade detection by the innate immune system in primary MDMs; they showed that disrupting either CA-CypA or CA-CPSF6 interaction caused the virus to trigger a type I IFN response and an antiviral state that prevented viral replication. In this study we have provided evidence to suggest that R132K/L136M reduces the ability of CA to interact with both CypA and CPSF6 and so we hypothesised that this mutant would trigger an IFN response in MDMs, like N74D. If so, this may mean that R132K/L136M has an even bigger infectivity defect *in vivo* than *in vitro*. Surprisingly, Rasaiyaah et al (unpublished observations) recently found that full-length HIV-1 (NL43-Bal) L136M, R132K/L136M or S41A/R132K/L136M replicated as efficiently as WT HIV-1 in primary MDMs and did not induce a detectable antiviral response. This could be interpreted to mean that R132K/L136M interacts sufficiently well with CypA and CPSF6 to be protected from innate immune detection but not to use the Nup358/TNPO3-dependent nuclear entry pathway. An alternative hypothesis is that the nuclear entry pathway of HIV-1 is different in HeLa and CD4+ T cells (in which R132K/L136M is defective) than it is in primary MDMs (in which R132K/L136M is not defective). Further characterisation of the nuclear entry pathway and the innate detection mechanism in MDMs is required before this observation can be fully understood. What this observation does suggest is that efficient replication of R132K/L136M HIV-1 in MDMs is not sufficient for the virus to achieve a high viral load *in vivo* but is sufficient for ongoing viral evolution (of S41A, or other compensatory mutations).

5 Chapter 5. Investigation and rational design of small-molecules interacting with the CPSF6-binding pocket of CA

5.1 Introduction

The majority of licensed ARVs target reverse transcriptase or protease, with some newer compounds targeting integrase or entry. One of the biggest problems facing these current treatments is the emergence of drug-resistant variants and so there is an increasing need for compounds inhibiting new therapeutic targets. CA is an exciting potential target for several reasons. Firstly, CA orchestrates both early and late events in HIV-1 infection and CA mutations have been shown to disrupt many stages of the viral lifecycle including reverse transcription, uncoating, nuclear entry, integration (and so potentially transcription and latency), assembly, and maturation. Furthermore, some regions of CA are very highly conserved; Li, et al. (407) recently reported a large-scale analysis of over 10,000 naturally occurring Gag sequences from across the 8 major HIV-1 subtypes in which they found 153 of the 231 CA residues to be completely conserved. They examined the binding sites of over 50 published Gag-binding inhibitors and found that the binding site of PF74, an antiviral compound that interacts with the same CA pocket as CPSF6 (111, 340), was the most highly conserved of all. Targeting a conserved site means that any resistance mutations that do arise are likely to have high fitness costs. It also means that the compound will have broad-spectrum activity across subtypes, the importance of which was exemplified by the development of Bevirimat, an assembly inhibitor that reached phase II clinical trials but proved ineffective against specific sub-types due to resistant polymorphisms in CA and SP1 (408).

The majority of published CA-targeting antivirals are described as assembly inhibitors (409). However, we are particularly interested in the recent discovery of CA-binding compounds that can inhibit early events of HIV-1 infection by disrupting the CA-CPSF6 interaction. PF-1385801 was identified in a high-throughput reporter-cell screen (410) for compounds inhibiting full-

length HIV-1 replication and analogs of this drug were designed in an attempt to improve potency. The most potent compound identified was PF-3450074 (PF74), the structure of which is shown in figure 47A. PF74 is reportedly active against a range of clinical HIV-1 isolates as well as laboratory strains (111). Serial passage of HIV-1 NL4-3 in MT2 cells in the presence of increasing concentrations of PF-1385801 lead to the outgrowth of a virus containing the CA mutation T107N (111, 411), which was later shown to escape inhibition by PF74 as well (412). CA residue T107 was also known to interact with CPSF6 and crystal structures of PF74 in complex with both monomeric CA_{NTD} and hexameric CA have now been solved, revealing that PF74 occupies the same binding pocket as CPSF6[68]$_{276-290}$ (111, 340, 342). As shown in figure 47B, PF74 interacts with two adjacent monomers within a CA hexamer, like CPSF6[68]$_{276-290}$. It hydrogen bonds with CA residues N57, Q63, and K70 of the first monomer and although it does not hydrogen bond with the second monomer it does interact with four residues (340). Comparison of the structures of CA in complex with CPSF6 peptide or PF74 reveals that a phenyl ring of PF74 almost exactly mimics the positioning of the phenyl ring of CPSF6 F284 (figure 47C). As one would predict, 100μM PF74 has been shown by ITC (isothermal titration calorimetry) to completely abolish CA-CPSF6[68]$_{276-290}$ interaction through direct competition (342).

A separate screen for inhibitors of VSV-G pseudotyped HIV-1 vector by Lamorte *et al.* (327) identified two further HIV-1 inhibitors that interact with CPSF6-binding pocket of CA, BI-1 and BI-2, which are both pyrrolopyrazolones and differ by a single hydroxyl group (the structure of BI-2 is shown in figure 47A). They found that whilst BI-2 is more potent than BI-1, it is also more cytotoxic, giving them similar therapeutic indices (TIs). Serial passage of NL4-3 in C8166 cells in the presence of increasing concentrations of BI-2 lead to the outgrowth of viruses containing the CA mutation T107N/A alongside A105T (327). Like T107, residue A105 interacts with CPSF6[68]$_{276-290}$ and crystal structures of BI-1 or 2 in complex with hexameric CA revealed that both compounds interact with the same binding pocket as CPSF6 and PF74 (327, 340, 342). BI-2 hydrogen bonds with residues N57, N74, and T107 within the first monomer
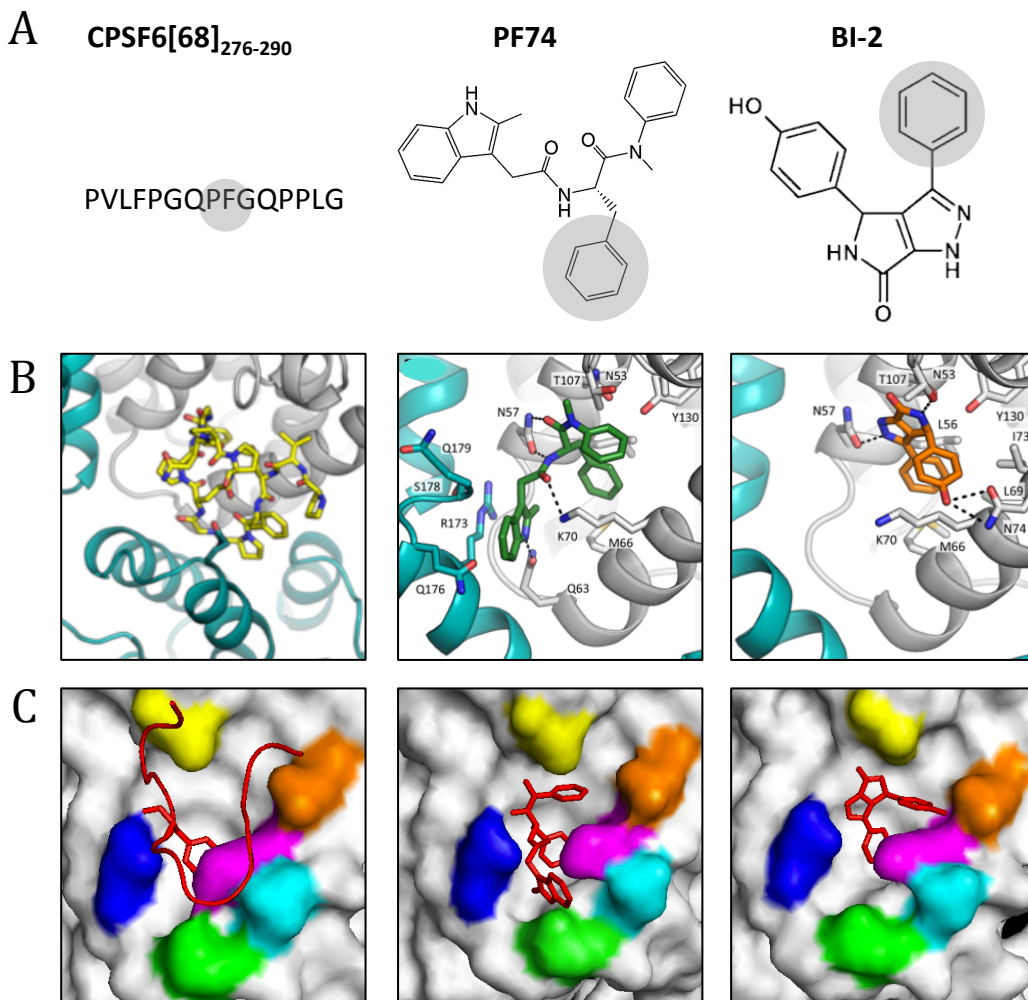
Figure 47. The structures of small molecule CPSF6 peptidomimetics
(A) Amino acid sequence of CPSF6[68]$_{276-290}$ or chemical structures of PF74 or BI-2. **(B)** Crystal structures of CPSF6[68]$_{276-290}$ in yellow, PF74 in green, or BI-2 in orange, in complex with hexameric CA in grey/teal. Images are taken from Price et al 2014. **(C)** Crystal structures of CPSF6[68]$_{276-290}$, PF74, or BI-2 in complex with hexameric CA, where only the first CA monomer is shown (PDBs 4U0A, 4U0E, and 4U0F, respectively, all from Price et al. 2014) . The CPSF6 peptide, PF74, and BI-1 are shown in red. CPSF6 residue F284 is highlighted with sticks to show that the phenyl group of this phenylalanine occupies the same pocket as phenyl groups of PF74 and BI-1. This residue and these phenyl groups are shaded in (A). To aid visualisation of CA structure CA residue N57 is in blue, Q63 in green, Q67 in cyan, K70 in pink, N74 in orange, and T107 in yellow, with the rest of CA in grey.

and places a phenyl ring in the same pocket as PF74 and CPSF6[68] F284 (figure 47C). However, whereas CPSF6 and PF74 also interact with the adjacent monomer of a hexamer, BI-2 does not (figure 47B).

Intriguingly, PF74 reportedly inhibits HIV-1 through a different mechanism than BI-1 and BI-2; whilst PF74 inhibits HIV-1 after entry but before reverse transcription (111, 113), BI-1 and 2 inhibit after reverse transcription but prior to 2-LTR circle formation (327). PF74 also purportedly inhibits late stages of HIV-1 infection; virus produced in the presence of PF74 contained similar amounts of p24 (as measured by ELISA) as virus produced in untreated cells, but the infectivity of these virions was reduced. EM of virions produced in cells treated with 7μM PF74 lack the dense conical cores that are typical of HIV-1 particles and exhibit a range of unusual morphologies (111), suggesting that PF74 disrupts the formation and/or maturation of cores. BI-1 is similarly been reported to disrupt CA-NC tube assembly, but only very weakly (327). Although the present study focuses on the early stages of infection, the fact that these compounds affect core assembly lead to the hypothesis that they inhibit early stages of infection by disrupting core structure.

As one would expect, both PF74 and BI-1 have been shown to compete CPSF6 from CA-NC tubes *in vitro* (345, 413). However, having shown that CPSF6 depletion and N74D mutation do not affect viral infectivity, we argue that these antiviral CPSF6 peptidomimetics cannot inhibit HIV-1 infection simply by competing endogenous CPSF6 off from CA. Therefore, these drugs must be having some direct effect upon the virus, as we have previously argued for antiviral forms of CPSF6 (discussed in section 3.3.2). PF74 and BI-1 have been tested for their effect upon core stability using the biochemical assays described in section 1.2.7.1 but, as with antiviral forms of CPSF6, the results have been inconsistent. Whilst one group has used the FoC assay to show that PF74 induces uncoating (113), another used it to show that PF74 prevents uncoating (90). Furthermore, whilst the FoC assay has been used to suggest that BI-1 (413) induces uncoating, the CA-NC tube assay has been used to show that it stabilises cores (327). In this study, we have sought to examine the mechanism

of PF74 and BI-1 inhibition using cellular assays in order to avoid the problems associated with biochemical assays.

It has recently been discovered that the pocket of CA bound by CPSF6, PF74, and BI-1 also interacts with Nup153 (340). It follows that PF74 has been shown to compete with $Nup153_{CTD}$ for CA-NC tube binding (83). It could therefore be suggested that PF74 and BI-1/2 inhibits HIV-1 by preventing interaction with Nup153. However, we have previously shown that CPSF6 depletion or N74D mutation causes HIV-1 to become independent of specific interaction with Nup153 and so we hypothesise that any drug that prevents CA-CPSF6 interaction would cause the virus to be insensitive to any drug inhibiting CA-Nup153 interaction.

Having found that HIV-1 triggers a potent innate immune response in the absence of CA-CPSF6 interaction, Rasaiyaah, et al. (92) tested whether HIV-1 infection of MDMs in the presence of PF74 lead to a type I IFN response. However, PF74 inhibited synthesis of viral cDNA, which is thought to be the PAMP detected by PRRs, and so did not trigger a detectable immune response (92). We therefore hypothesise that a drug that competes CPSF6 from CA but does not inhibit reverse transcription, such as BI-1/2, may 'uncloak' the virus and trigger an innate response. However, these compounds were toxic to primary MDMs at doses that inhibit the virus (Rasaiyaah et al unpublished observations). If such a drug could be designed, it would potentially have use as an HIV-1 therapeutic and also as a prophylactic; by inducing an antiviral state at the site of primary infection it may prevent spread of the virus. Regardless of whether it induced an innate immune response, any antiviral compound binding to the same CA pocket as CPSF6 is of great interest because we predict that escape mutations in CA that prevent the drug from binding but would also prevent CPSF6 from binding, and these mutants would therefore induce an antiviral response.

Shi, et al. (113) recently reported that CypA depletion, G89V or P90A mutation, or CsA reduced the efficacy of PF74. If CypA and PF74 stabilise cores in a

cooperative manner (and PF74 inhibits because its stabilisation is not reversible), then CsA may lead to a reduction in PF74 efficacy. Alternatively, if CA-CypA interaction causes conformational changes that encourage CA to interact with PF74, CsA may reduce the affinity of CA for PF74. Ambrose, et al. (358) reported that N74D mutation reduces the amount of CypA that co-immunoprecipitates with CA, which suggests that CA-CPSF6 interaction could alter the affinity of CA for CypA interaction. Together, these reports point to cooperative binding between CA, CypA, and CPSF6. Consistent with this, CypA and CPSF6 play very similar roles in HIV-1 infection (discussed in section 3.3.5).

In this study, we have further probed the mechanism/s of action of PF74 and BI-1. In order to avoid the previously discussed caveats of the biochemical assays used in many previous studies, we focused on using cellular assays and measuring infectivity as a read-out. Due to the fact that the mechanism of action of CPSF6, PF74, or BI-1/2 is not fully understood, we could not easily rationally design drugs to have a specific effect upon the virus. Instead we sought to identify compounds with novel phenotypes and to use these as tools to learn more about the effects of ligands interacting with this conserved CA pocket. We did, however, design compounds with the general aims of making more extensive contacts with CA or disrupting the structure of viral core.

5.2 Results

5.2.1 HA-hCPSF6[68]-358, PF74 and BI-1 exhibit different inhibition profiles despite interacting with the same pocket of CA

HA-hCPSF6[68]-358, PF74 and BI-1 are all reported to interact with the same pocket of HIV-1 CA and to inhibit the early stages of infection. We therefore sought to compare their inhibition profiles. We found that HA-hCPSF6[68]-358 inhibits infectivity, reverse transcription, and 2-LTR-circle formation to similar magnitudes of 21 to 30-fold (figures 48A, B, and C, respectively). When titrating PF74 or BI-1 we began to observe cell death at 30μM PF74 or 90μM BI-1, respectively, in agreement with previously published CC_{50} values (111, 327). We therefore maximally used 20μM PF74 or 80μM BI-1 in the following experiments. Upon titrating PF74, we found that it inhibits HIV-1 infectivity in two phases; by 0.5μM it inhibits infectivity approximately 10-fold but then plateaus until ~4μM, after which the level of inhibition increases again (figure 48D and G). This biphasic pattern of inhibition could suggest that PF74 inhibits via two different mechanisms that act at distinct concentrations. Blair *et al.* (111) reported that PF74 inhibits infection prior to reverse transcription. However, having titrated this drug we found that it only inhibits reverse transcription at concentrations above ~5μM and the degree of the block to reverse transcription never accounts for the degree of the block to infectivity (figure 48E). Below 5μM, PF74 inhibits after reverse transcription but prior to 2-LTR-circle formation (figure 48F). We noted that inhibition of reverse transcription begins at approximately the same concentration as the second phase of the infectivity inhibition curve (4-5μM). If, as suggested earlier, PF74 inhibits via two different mechanisms at different concentrations, this observation would suggest that only the second mechanism inhibits reverse transcription.

In contrast to PF74, BI-1 exhibits a single phase of inhibition (figure 48H). In agreement with a previous report (327) we found that BI-1 inhibits after reverse transcription but prior to nuclear entry (figures 48I and J, respectively). This could be interpreted to mean that PF74 and BI-1 inhibit HIV-1 by entirely
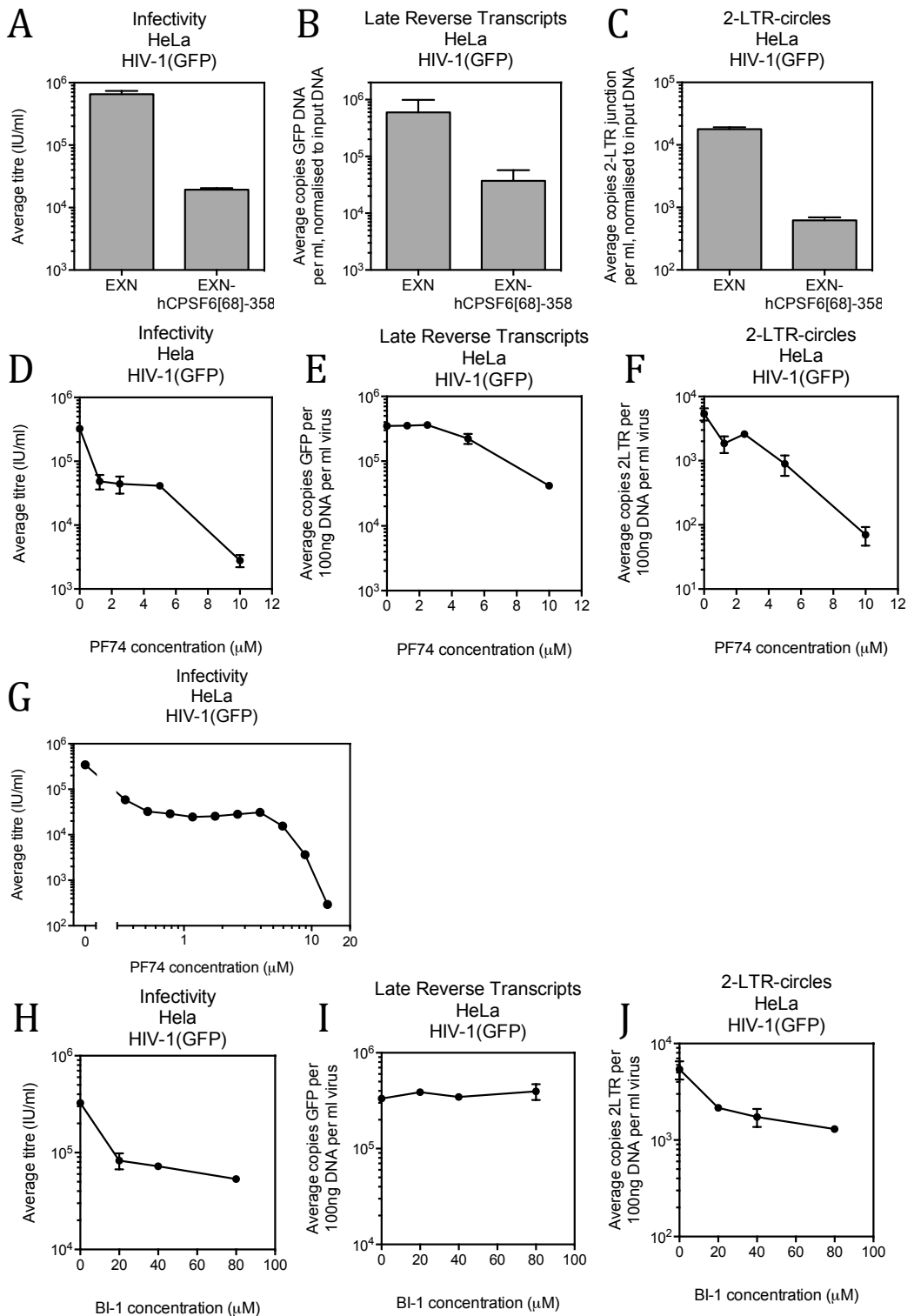
Figure 48. The inhibition profiles of hCPSF6[68]-358, PF74, and BI-1
(A-C) HeLa cells stably expressing hCPSF6[68]-358 or empty vector were infected with a titration of WT HIV-1(GFP). **(D-F)** Unmodified HeLa cells were infected with a titration of WT HIV-1(GFP) in the presence of a titration of PF74. (D) Viral titre was measured at 48hrs p.i. by flow cytometry. (E) Late reverse transcripts were measured 6hrs p.i. by qPCR. (F) 2-LTR circles were measured at 18hrs p.i. by qPCR. **(G)** As in (D) but a larger titration of PF74. **(H-J)** as in (D-F) but using BI-1. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

220

different mechanisms. Alternatively, it could mean that BI-1 inhibits by the same mechanism as low-dose PF74. Importantly, BI-1 may also exhibit a biphasic inhibition curve if it could be used at higher doses without being toxic.

5.2.2 HIV-1 loses sensitivity to Nup358/TNPO3 depletion in the presence of PF74/BI-1

Earlier in this study we showed that CA-CPSF6 interaction commits HIV-1 to using a particular nuclear entry pathway, which is dependent upon Nup358, TNPO3, and Nup153 (figure 23). PF74 and BI-1 have been shown to compete with CPSF6 for CA-NC binding *in vitro* (345, 413) and it follows that these antiviral compounds ought to also prevent CA from interacting with endogenous CPSF6 in the context of an infection. We therefore hypothesised that PF74 and BI-1 prevent the virus from using the CPSF6-dependent nuclear entry pathway. We infected HeLa cells stably depleted of TNPO3 or Nup358 in the presence of a titration of PF74 or BI-1. We found that the degree of inhibition caused by PF74 negatively correlated with the degree of inhibition caused by TNPO3/Nup358 depletion. In other words, increasing concentrations of PF74 caused the virus to become decreasingly sensitive to Nup358/TNPO3 depletion (raw/fold data in figure 49A/B). We also observed a similar trend for BI-1 (raw/fold data in figure 49C/D). The fact that PF74 and depletion of Nup358/TNPO3 are not additive in their ability to inhibit HIV-1 suggests that they inhibit the same step of HIV-1 infection, nuclear entry. Importantly, we could not find a concentration of PF74/BI-1 that reduced sensitivity to Nup358/TNPO3 depletion without also inhibiting infection. This suggests that the same (or lower) PF74/BI-1 concentration is required to inhibit the virus as is required to free the virus from commitment to the TNPO3/Nup358-dependent nuclear entry pathway.

5.2.3 Inhibition by PF74, BI-1, or HA-hCPSF6[68]-358 is reversible

As discussed in section 5.1, biochemical assays used to test the effect of CPSF6 truncations, PF74, or BI-1 upon core stability have produced inconsistent results. Some studies have suggested that these inhibitors work by destabilising
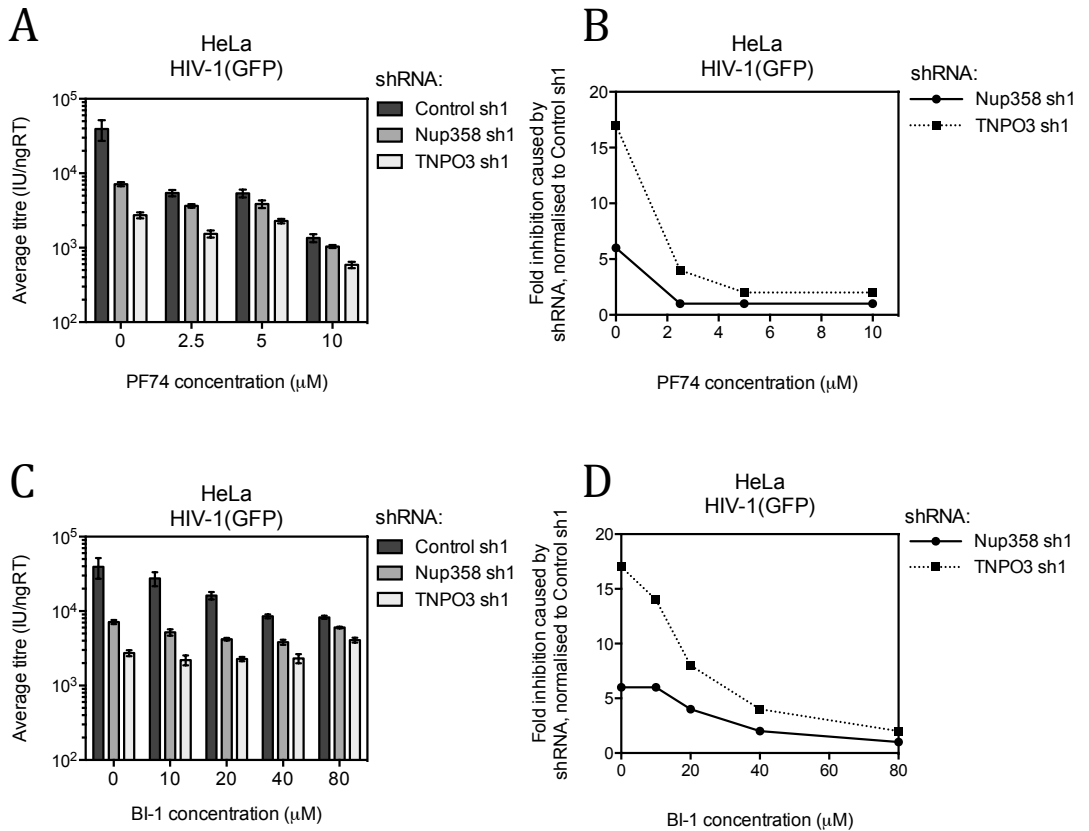
Figure 49. PF74 and BI-1 cause HIV-1 to become less sensitive to Nup358 and TNPO3 depletion
HeLa cell clones stably expressing Nup358- or TNPO3-targeting shRNA or a non-targeting control shRNA were infected with a titration of WT HIV-1(GFP) in the presence of a titration of **(A)** PF74 or **(C)** BI-1. **(B)** and **(D)** show data in (A) and (C), respectively, as fold difference between titre in control cells and titre in Nup358/TNPO3-depleted cells. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

and prematurely uncoating cores (113, 344, 413). If this were true, then inhibition would be irreversible. We tested whether inhibition by PF74 is reversible by performing a washout assay, as depicted in figure 50A. We infected HeLa cells in the presence of a titration of PF74 and allowed infection to proceed for 6 hours, a time point at which we had previously shown PF74 to be effective in inhibiting reverse transcription (figure 48E). We then removed the infected cells' media and replenished it, either with drug-containing media (wash) or with unsupplemented media (washout). We repeated this 5 times at 5min intervals to ensure efficient wash out of the drug. We then allowed infection to continue as normal and assessed infectivity by flow cytometry at 48 hours post-infection. We found that inhibition by 5 or 10μM PF74 was almost completely reversible by washout (figure 50B). Presuming that uncoating is an irreversible process, this observation strongly argues against the notion that PF74 induces premature uncoating. At 20μM PF74, we found that infectivity could only be partially rescued. This could simply be because wash out of the drug was incomplete. However, an alternative explanation is that the second 'phase' of PF74-mediated inhibition, which only occurs at high concentrations and inhibits reverse transcription, is caused by the drug uncoating the virus. We therefore sought to test whether 20μM PF74-mediated inhibition of reverse transcription was reversible. We performed a washout as described previously, but also measured late reverse transcripts by qPCR at the time of washout (6hrs post-infection) and 6hrs after the washout (12hrs post-infection), as depicted in figure 50C. This ensured that any viral cores that were inhibited for reverse transcription at 6hrs post-infection had another 6hrs to produce reverse transcripts after the drug had been removed. We found that 20μM PF74 inhibited reverse transcription 8-fold at 6 hours post-infection, as expected (figure 50D). 6 hours after washing (12 hours post-infection) reverse transcription was inhibited 11-fold in 'wash' cells but only 2.5-fold in 'washout' cells (figure 50E). This shows that PF74-mediated inhibition of reverse transcription is reversible. This argues against the hypothesis that infection can only be partially rescued at 20μM PF74 because the second phase of inhibition inhibits reverse transcription by inducing uncoating.
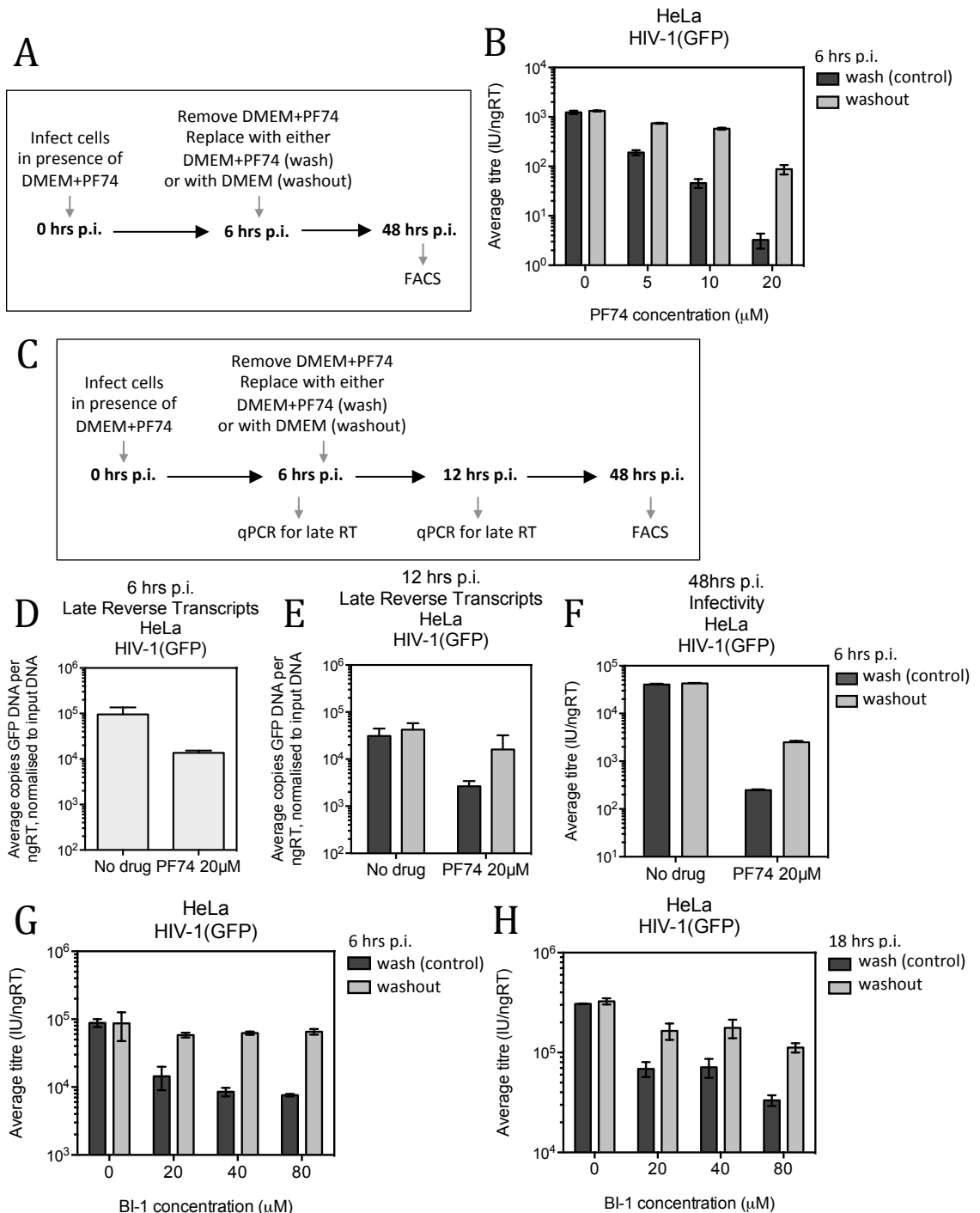
Figure 50. Inhibition of HIV-1 by PF74 or BI-1 is reversible by washout
(A) Diagram of the experimental procedure carried out in (B). **(B)** As depicted in (A), HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74. 6hrs p.i. media was replaced 5 times, either with more PF74-containing media ('wash') or with unsupplemented media ('washout'). Viral titre was measured 48hrs later by flow cytometry. **(C)** Diagram of the experimental procedure carried out in (D-F). As depicted in (C), HeLa cells were infected with a titration of HIV-1(GFP) in the presence of either 20uM PF74 or DMSO. 6hrs p.i. media was replaced 5 times, either with more PF74-containing media ('wash') or with unsupplemented media ('washout'). Late reverse transcripts were measured **(D)** 6hrs p.i. or **(E)** 12hrs p.i. by qPCR. **(F)** Viral titre was measured 48hrs after wash/washout by flow cytometry. **(G)** As in (B) but used a titration of BI-1. **(H)** As in (G) but carried out wash/washout at 18hrs p.i. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

224

We similarly tested whether BI-1-mediated inhibition of infectivity could be rescued by washout. We found that washout of BI-1 at 6 hours post-infection completely rescued infectivity (figure 50G). We assume that BI-1 interacts with viral cores at a similar stage of infection as PF74 but, given that BI-1 does not inhibit reverse transcription, we cannot measure whether BI-1 is having any effect upon the virus at this time point. We therefore also washed out BI-1 at 18 hours post-infection, a time point at which we knew that BI-1 was effective in inhibiting 2-LTR-circle formation (figure 48J). Washout of BI-1 at 18 hours post-infection partially rescued HIV-1 infectivity (figure 50H). We propose that rescue is partial because viral complexes that are stalled in the cytoplasm may begin to be degraded by 18 hours post-infection.

Of course, we cannot use a washout assay to measure whether HA-hCPSF6[68]-358-mediated inhibition is reversible. Instead, we tested whether HA-hCPSF6[68]-358-mediated inhibition of reverse transcription is reversible by displacing this protein from cores using PF74 or BI-1 at concentrations that we had previously shown do not inhibit reverse transcription (figures 48E and I, respectively). We infected HA-hCPSF6[68]-358 expressing HeLa cells or control cells with WT HIV-1 and allowed infection to progress for 6 hours, a time point at which we have previously shown HA-hCPSF6[68]-358 to inhibit reverse transcription (figure 48B). We then supplemented the infected cells' media with either 80μM BI-1 or 5μM PF74. We then allowed infection to progress for a further 6 hours, measuring reverse transcripts every 2 hours. In control cells, addition of PF74/BI-1 at 6 hours post-infection did not affect reverse transcription (figure 51 A/C). However, in HA-hCPSF6[68]-358-expressing cells, reverse transcription was inhibited at 6 hours post-infection and addition of PF74/BI-1 lead to an increase in the number of reverse transcripts over the following 6 hours (figures 51 B/D). These data confirm that this C-terminal truncation of CPSF6 can be displaced by 5μM PF74 or 80μM BI-1 in the context of an infection. They also reveal that, like PF74, HA-hCPSF6[68]-358 inhibits reverse transcription in a reversible manner, arguing that this restriction factor
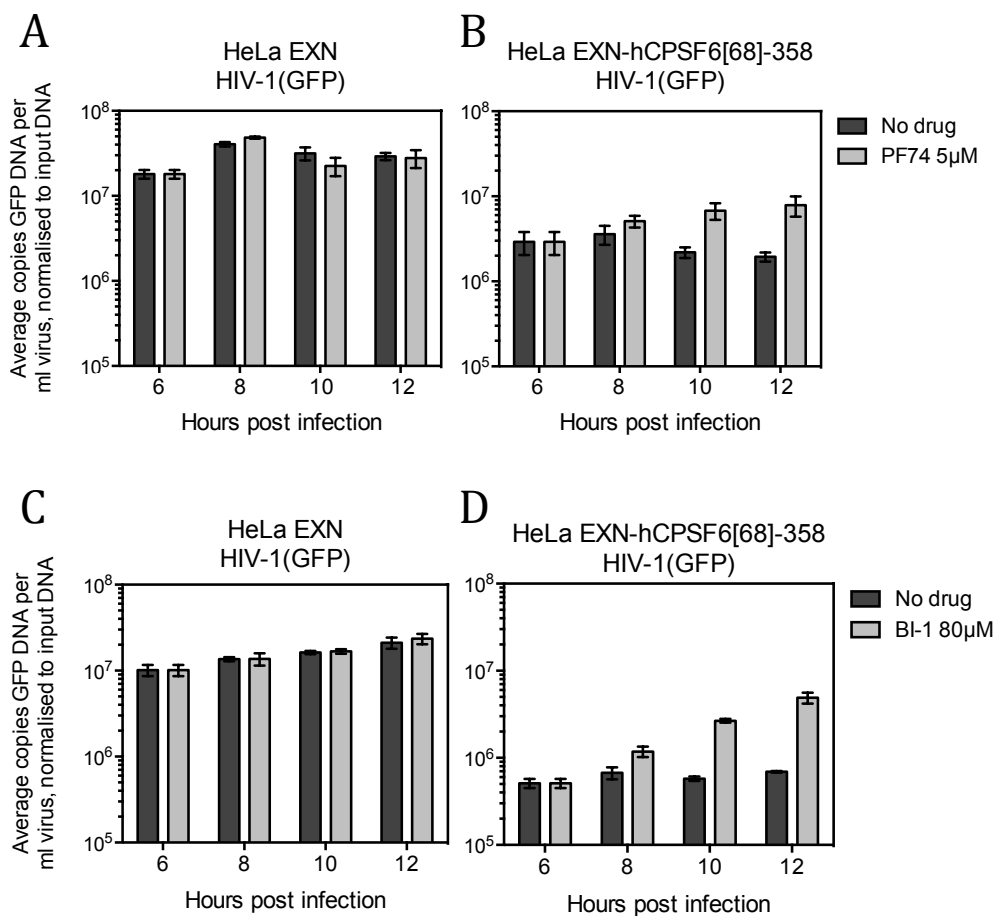
Figure 51. Inhibition of HIV-1 reverse transcription by hCPSF6[68]-358 is reversible by addition of PF74 or BI-1

HeLa cells stably expressing **(A/C)** empty vector or **(B/D)** hCPSF6[68]-358 were infected with a titration of HIV-1(GFP). At 6hrs p.i. media was supplemented with **(A/B)** 5μM PF74 or **(C/D)** 80μM BI-1. **(A-D)** Late reverse transcripts were measured at 6, 8, 10, or 12hrs p.i. by qPCR. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

does not prematurely uncoat HIV-1 cores.

5.2.4 PF74 and BI-1 inhibit HIV-1 at a lower occupancy than HA-hCPSF6[68]-358

The observations made in figure 51 revealed that HA-hCPSF6[68]-358, which inhibits reverse transcription, can be displaced by concentrations of PF74 or BI-1 that do not inhibit reverse transcription. This suggested that HA-hCPSF6[68]-358 may inhibit reverse transcription at a lower CA occupancy than PF74/BI-1, which could suggest that they inhibit HIV-1 via different mechanisms. To investigate this possibility further, we infected HA-hCPSF6[68]-358-expressing cells in the presence of a titration of PF74 or BI-1. We were intrigued to find that PF74 can rescue HIV-1 from inhibition by HA-hCPSF6[68]-358. For example, HA-hCPSF6[68]-358 expression inhibited HIV-1 infection 44-fold but, together, HA-hCPSF6[68]-358 and 5μM PF74 only inhibited infection by 6.5 fold (figure 52A). In other words, 5μM PF74 rescued infectivity from HA-hCPSF6[68]-358 by 7-fold. We observed the same pattern when measuring reverse transcription; HA-hCPSF6[68]-358 expression inhibited reverse transcription 11-fold but, together, HA-hCPSF6[68]-358 and 2.5μM PF74 only inhibited reverse transcription 3-fold (figure 52B). In other words, 2.5μM PF74 rescued reverse transcription by 3.5-fold. Similarly, we found that 80μM BI-1 was able to rescue HIV-1 infectivity from HA-hCPSF6[68]-358 by 2-fold (figure 52C) and reverse transcription by 5-fold (figure 52D). These observations inform us that PF74 and BI-1 are able to displace HA-hCPSF6[68]-358 from CA at concentrations at which they are less inhibitory than HA-hCPSF6[68]-358. We interpret this to mean that HA-hCPSF6[68]-358 inhibits HIV-1 (infectivity and reverse transcription) at a lower CA occupancy than PF74.
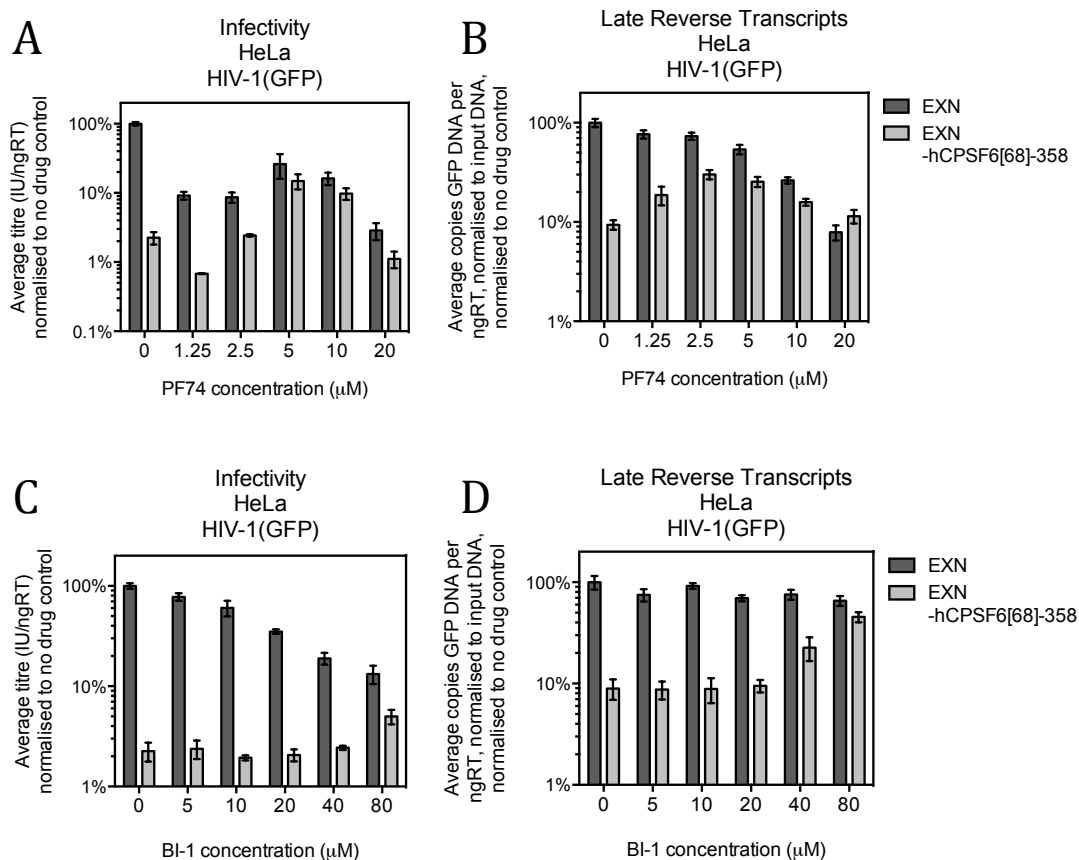
Figure 52. PF74 and BI-1 inhibit HIV-1 at a lower CA occupancy than hCPSF6[68]-358

HeLa cells stably expressing hCPSF6[68]-358 or empty vector were infected with a titration of HIV-1(GFP) in the presence of a titration of **(A-B)** PF74 or **(C-D)** BI-1. **(A/C)** Viral titre was measured 48hrs p.i. by flow cytometry. **(B/D)** Late reverse transcripts were measured at 6hrs p.i. by qPCR. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

5.2.5 Compounds disrupting CA-CPSF6 interaction and compounds disrupting CA-CypA interaction can be antagonistic or synergistic in their ability to inhibit HIV-1 infection, in a dose-dependent manner

Shi et al. reported (113) that CsA antagonises the antiviral activity of PF74 in HeLa cells. Hilditch *et al.* (unpublished observations) similarly found that CsA partially rescues HIV-1 infectivity from HA-hCPSF6[68]-358-mediated inhibition in HeLa cells. These observations suggest that CypA and CPSF6 may act cooperatively. In order to investigate this relationship further, we tested the effect of SmBz (a non-immunosuppressive CsA analog) upon PF74 or BI-1. In HeLa cells, 10μM SmBz had no significant affect upon viral infectivity (figure 53A) or reverse transcription (figure 53B). Despite this, SmBz did have a significant effect upon the ability of PF74 to inhibit HIV-1 infection (figure 53C). 10μM SmBz antagonised PF74 concentrations below ~2μM, rescuing infectivity by up to 5.5-fold. However, SmBz was synergistic with higher concentrations of PF74, enhancing the inhibitory effect of PF74 by up to 6-fold. The overall result of these concentration-dependent effects is that the biphasic nature of the PF74 inhibition curve is lost in the presence of SmBz, such that there is no longer a plateau. Interestingly, although SmBz enhanced the ability of PF74 to inhibit infection, it did not affect the ability of PF74 to inhibit HIV-1 reverse transcription (figure 53D).

We next examined the relationship between these compounds in a cell type in which SmBz does inhibit HIV-1 infectivity. In TE671s, 10μM SmBz inhibited HIV-1 infection 5-fold (figure 54A) and reverse transcription 4-fold (figure 54B). As in HeLa cells, the effect of SmBz on PF74 varied with PF74 concentration in TE671s (figure 54C). However, SmBz did not antagonise PF74 in TE671s. Instead, the inhibitory effects of SmBz and PF74 were either additive or synergistic. For example, 4μM PF74 and 10μM SmBz inhibited HIV-1 infection 9- and 5-fold, respectively, but in combination they inhibited the virus 118-fold. Interestingly, when measuring inhibition of reverse transcription, the effects of SmBz and PF74 were only additive, not synergistic (figure 54D).
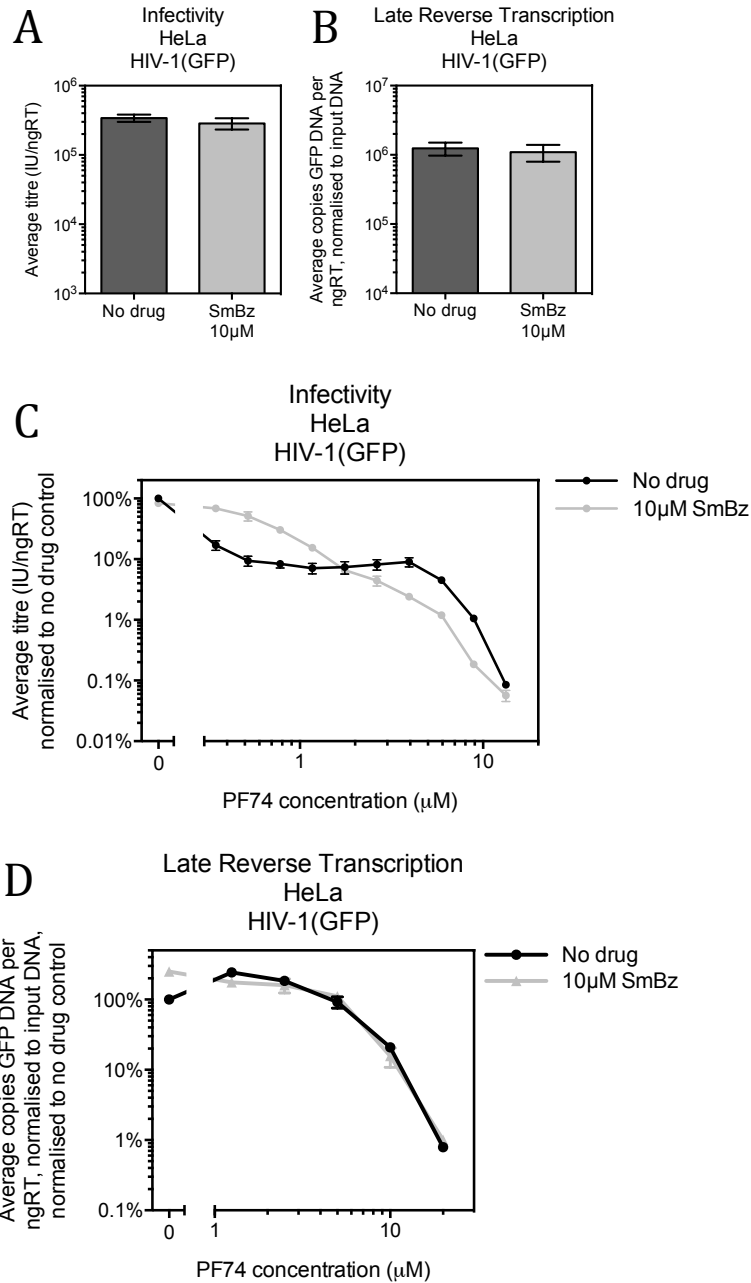
Figure 53. The relationship between PF74 and SmBz in HeLa cells
(A-B) HeLa cells were infected with a titration of HIV-1(GFP) +/- 10μM SmBz. **(C-D)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74 +/- 10μM SmBz. **(A/C)** Viral titre was measured 48hrs p.i. by flow cytometry. **(B/D)** Late reverse transcripts were measured 6hrs p.i. by qPCR. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).
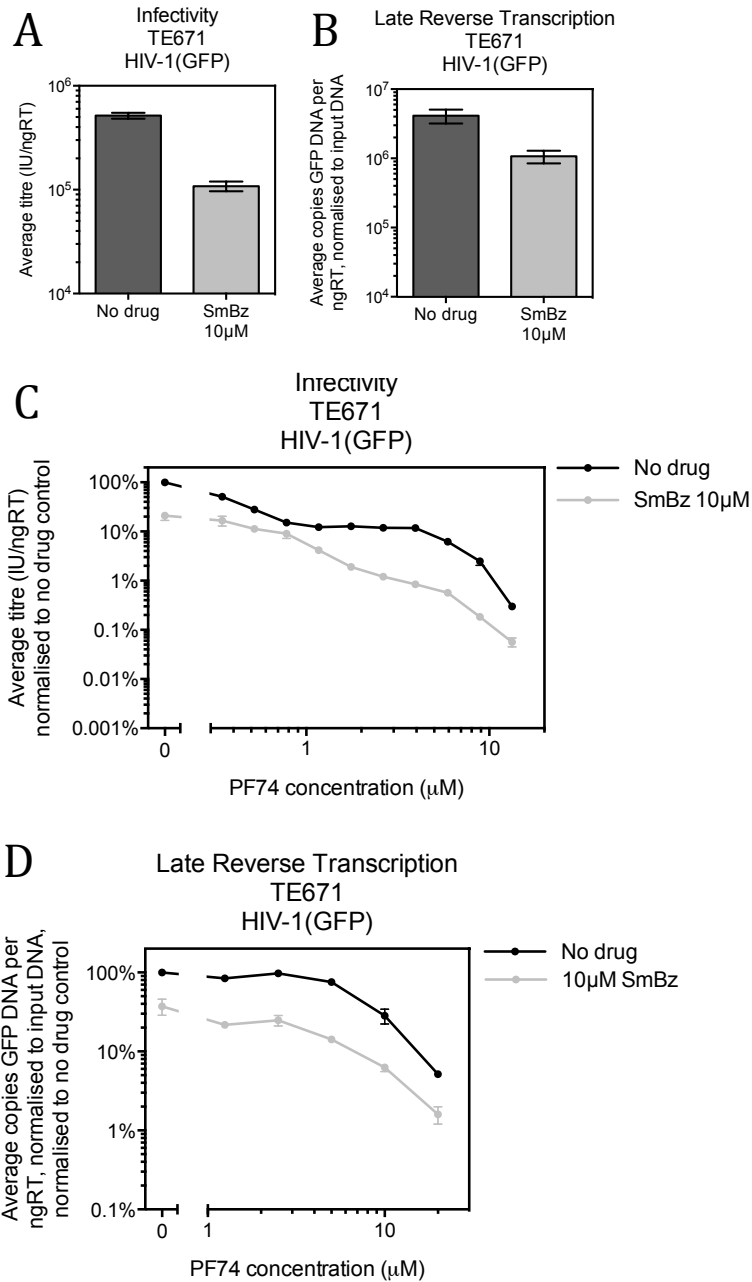
Figure 54. The relationship between PF74 and SmBz in TE671cells
(A-B) TE671 cells were infected with a titration of HIV-1(GFP) +/- 10μM SmBz. **(C-D)** TE671 cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74 +/- 10μM SmBz. **(A/C)** Viral titre was measured 48hrs p.i. by flow cytometry. **(B/D)** Late reverse transcripts were measured 6hrs p.i. by qPCR. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

We found that SmBz had a similar relationship with BI-1 as with low concentrations of PF74. In HeLa cells, SmBz antagonised BI-1, rescuing infectivity by up to 7-fold (figure 55A) whilst in TE671s, SmBz and BI-1 were additive (figure 55B). This is consistent with the hypothesis that BI-1 and low-dose PF74 inhibit HIV-1 by the same mechanism.

5.2.6 Rational design of small molecules to fit the CPSF6-binding pocket of CA

In this study we have provided evidence that PF74 and BI-1 displace CPSF6, inhibit the Nup358/TNPO3-dependent nuclear entry pathway, inhibit in a reversible manner (and so do not induce uncoating), and require a higher CA occupancy than HA-hCPSF6[68]-358 to inhibit HIV-1. Whilst we, and others, hypothesise that these compounds stabilise cores, the actual mechanism of inhibition remains disputed. What is clear is that PF74 and BI-1 interact directly with HIV-1 CA and displace CPSF6, but also have some other direct effect upon the viral core.

In this study, we designed a series of small molecules that were based upon the structure of PF74 and so were predicted to interact with the CPSF6-binding pocket of CA. Our aim was to design molecules that have stronger and/or different phenotypes from PF74/BI-1 and then to compare the effect of these molecules upon CA structure, such that we can better understand the effect of changes in core structure upon the early stages of viral infection. Whilst this was primarily an exploratory study, we hypothesised that compounds making more extensive contacts with CA may be more potent antivirals, particularly if they mimic HA-hCPSF6[68]-358 more closely and thus inhibit at a lower CA occupancy. We were also interested in whether the compounds we created would inhibit reverse transcription, given that PF74 does but BI-1 does not. Furthermore, although we have shown that PF74 and BI-1 do not induce uncoating, we hypothesised that larger molecules binding to this pocket might disrupt core structure and induce uncoating. We hypothesise that such a drug would not inhibit the virus, but would cause WT HIV-1 to phenocopy N74D and

Infectivity
HeLa
HIV-1(GFP)

Infectivity
TE671
HIV-1(GFP)
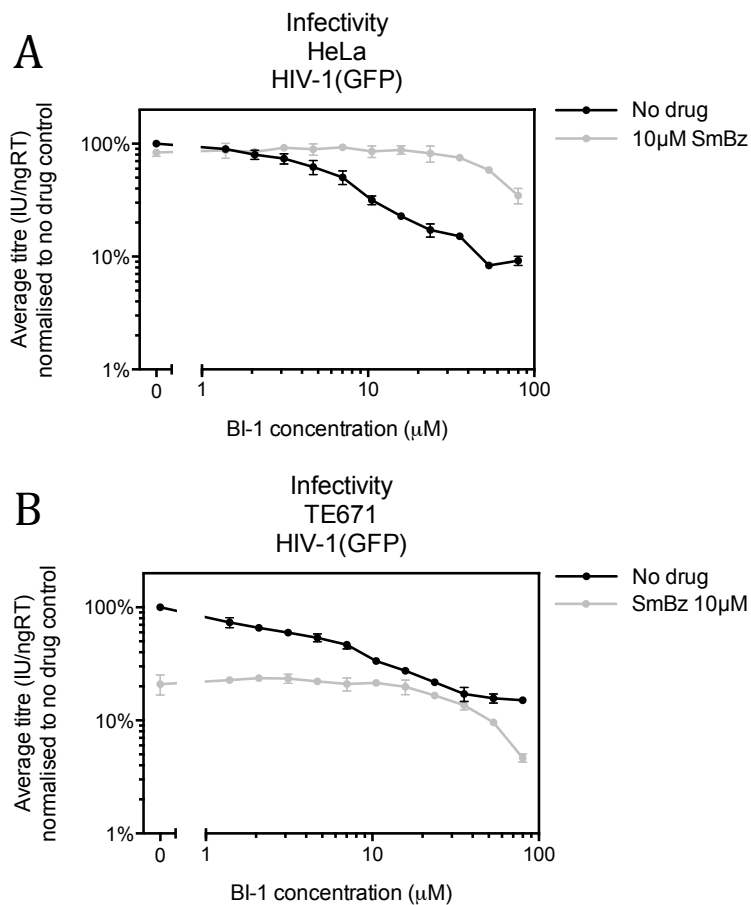
Figure 55. The relationship between BI-1and SmBz in HeLa or TE671 cells
(A) HeLa or **(B)** TE671 cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74 +/- 10μM SmBz. Viral titre was measured 48hrs p.i. by flow cytometry. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

infect via a Nup358 and TNPO3-independent pathway. Furthermore, we hypothesise that it would trigger an innate immune response in MDMs, like SmBz (92).

All SAL compounds used in this study were designed, modeled, and synthesised by Dave Selwood, Justin Warne, Edith Chan, Richard Angell, and Sally Oxenford.

5.2.6.1 CPSF6 peptidomimetics differ in their inhibition profiles

As shown in figure 56, we divided PF74 into four chemical regions– the indole acetic acid N-terminal cap, the phenyl group (which mimics CPSF6[68] residue F283), the core scaffold, and the C-terminal cap – and explored modifications of each region in turn. The crystal structures of PF74 in complex with CA all show that the phenyl group, which mimics CPSF6[68] F284, fits its CA binding pocket very tightly (340-342). We therefore predicted that addition of other groups around this ring would occlude binding and so we have not attempted to modify this group as yet. In contrast, the indole group is the region of PF74 that interacts with the second monomer in hexameric CA (figure 47B). We therefore tested the effect of substituting this indole group for different functional groups that may interact differently with this second monomer. However, replacing the indole for a carboxyl group (SAL-1 and 2) or a carboxamide group (SAL-3) as shown in figure 57A resulted in a significant loss of antiviral activity as compared to PF74 (figure 57B).

The core scaffold of PF74 comprises an amide bond that is primarily in the *trans* conformation in solution but is fixed in the *cis* conformation when in complex with CA. By cyclising this amide group to form a triazole ring, we fixed the compound in the conformation it adopts when bound to CA. We first created SAL-4 (figure 58A), which has a thiol group attached to the triazole ring. A crystal structure of SAL-4 bound to hexameric CA (provided by James L.), shown in figure 58B, confirmed that it interacts with CA in an extremely similar conformation to PF74, hydrogen bonding with the same residues (N57, Q63, and K70). We found SAL-4 to be significantly more potent than BI-1 but less
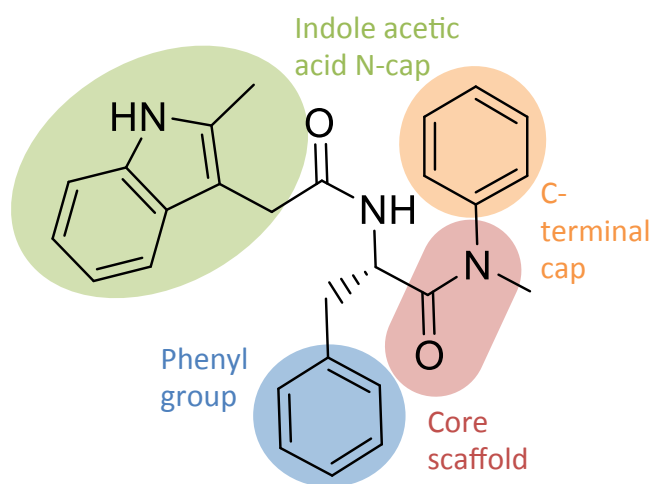
Figure 56. The chemical regions of PF74
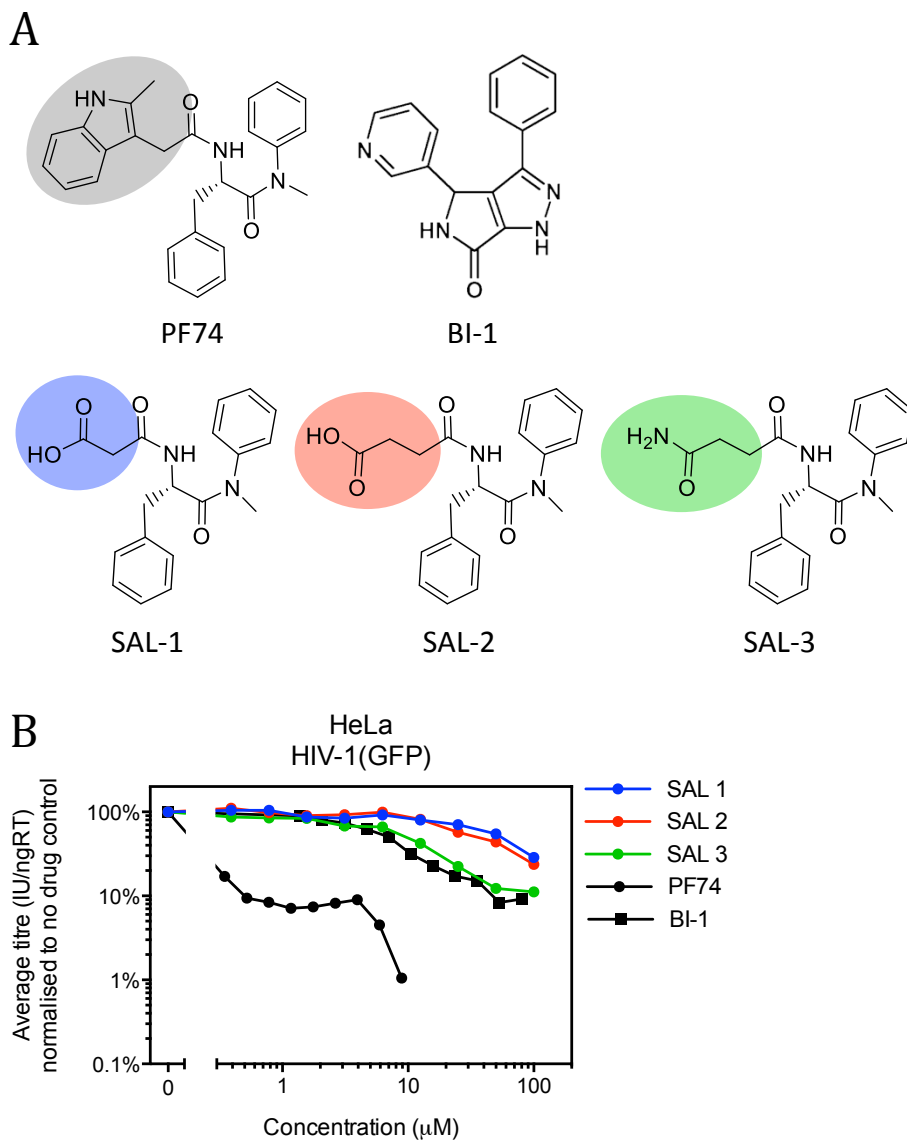
Figure 57. The effects of replacing the indole group of PF74
(A) Chemical structures of PF74, BI-1, SAL-1, 2, and 3. The shaded ovals highlight regions that differ between compounds. **(B)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of the compounds depicted in (A). All data are representative of at least 2 independent experiments.

A

PF74 *trans*   PF74 *cis*   SAL-4

B

CA$_{Hex}$ and PF74   CA$_{Hex}$ and SAL-4

C

HeLa
HIV-1(GFP)

- SAL 4
- BI-1
- PF74

Figure 58. The effects of cyclising the core scaffold of PF74
(A) Chemical structures of PF74 (in *trans* or *cis*) and SAL-4. The shaded ovals highlight regions that differ between compounds. **(B)** A comparison of crystal structures of PF74 (PDB 4U0E from Price et al. 2014) or SAL-4 in complex with hexameric CA. Peptides shown in red, CA in grey, **(C)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74, BI-1, or SAL-4. All infectivity data are representative of at least 2 independent experiments.

237

than PF74 (figure 58C). We also created other cyclised compounds in which the thiol group of SAL-4 was replaced by a thiomethyl group (SAL-15), a hydroxyl group (SAL-25), or much larger groups (SAL-24 and 147), with the aim of making up further interactions between CA and the drug (figure 59A). Whilst SAL-25 and 147 were both weakly antiviral, SAL-15 and 24 both had a similar potency as SAL-4 (figure 59B). However, whilst SAL-4 had a normal inhibition curve, SAL-15 and 24 exhibited the biphasic inhibition curve typical of PF74. We have previously discussed the possibility that BI-1 inhibits HIV-1 via the same mechanism as PF74 but simply never reaches this level of inhibition due to its lower affinity and/or solubility. However, our observation that molecules with very similar potencies can exhibit these two different phenotypes strongly argues against this idea. We have also previously discussed the possibility that PF74 has a biphasic inhibition curve due to its ability to interact with two adjacent monomers, which BI-1 does not. However, we have shown that SAL-4 interacts with two adjacent monomers and yet does not show the same pattern of inhibition as PF74.

We also tested the effect of altering the indole group of SAL-4. We found that methylating the nitrogen of the indole (SAL-13) or adding a chloro group to the indole (SAL-19) reduced potency, whereas adding a methoxy group (SAL-14) increased toxicity without improving potency (figures 60A and B). We also found that demethylating the indole of SAL-4 (SAL-135), adding a CH_2 linker (SAL-157), or replacing the indole with a benzoindole group (SAL-151) diminished antiviral activity (figures 61A andB). Intriguingly, adding a carbonyl group to the indole to form an oxindole (SAL-143) created a compound that actually increased HIV-1 infectivity by 2-fold.

When comparing the structure of CPSF6[68]$_{276-290}$ in complex with CA with that of PF74/SAL-4, we noted that CPSF6[68] residues V277 and L278 make contacts with CA in a pocket not reached by PF74/SAL-4 (figure 62A). With the aim of designing a compound that mimics CPSF6 more closely, we attempted to reach this pocket by extending the C-terminal cap phenyl group of SAL-4 in a variety of ways. Adding small methyl, methoxy, or tertbutyl groups to this
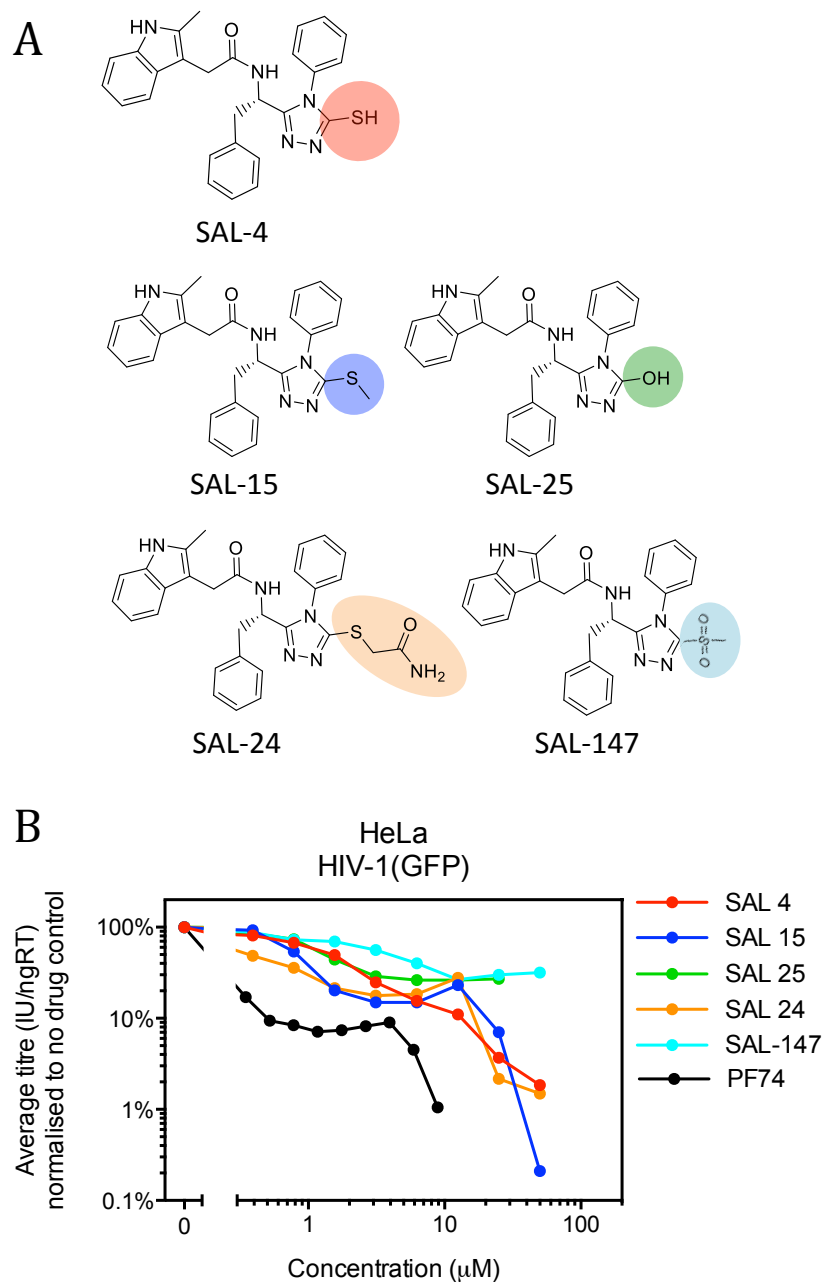
Figure 59. The effects of modifying the core scaffold of SAL-4
(A) Chemical structures of SAL-4, 15, 25, 24, and 147. The shaded ovals highlight regions that differ between compounds. **(B)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74 or the compounds depicted in (A). All data are representative of at least 2 independent experiments.
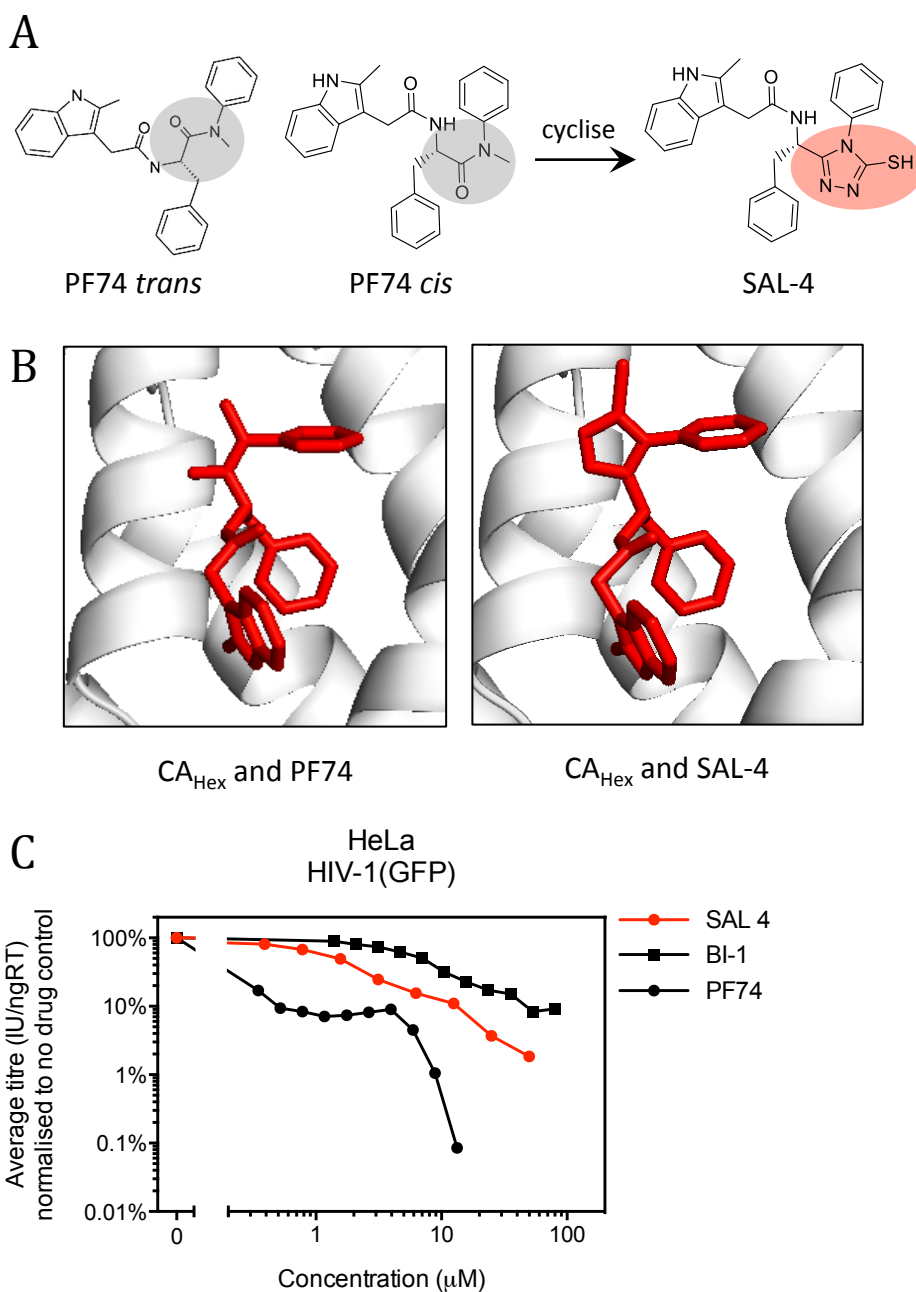
Figure 60. The effects of modifying the indole group of SAL-4
(A) Chemical structures of SAL-4, 13, 19, and 14. **(B)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74 or the compounds depicted in (A). The shaded ovals highlight regions that differ between compounds. All data are representative of at least 2 independent experiments.
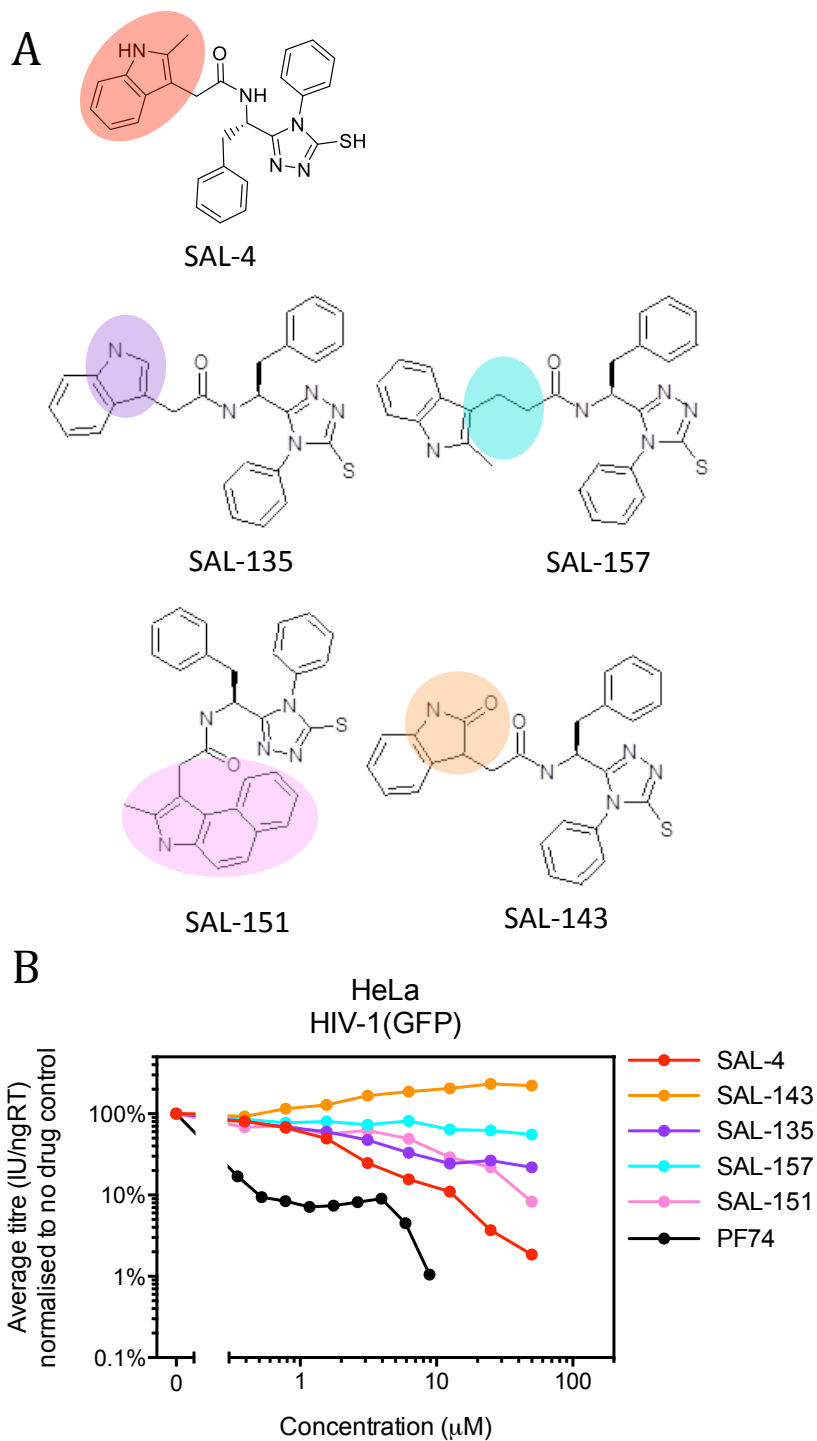
Figure 61. The effects of modifying the indole group of SAL-4 continued
(A) Chemical structures of SAL-4, 135, 157, 151, and 143. **(B)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74 or the compounds depicted in (A). The shaded ovals highlight regions that differ between compounds. All data are representative of at least 2 independent experiments.

A

CA_{Hex} and CPSF6[68]$_{276-290}$
Residues F284, V277, and
L278 in bold

CA_{Hex} and SAL-4
Arrow shows pocket

B

SAL-4

SAL-11

SAL-17

SAL-12

SAL-10

C

HeLa
HIV-1(GFP)

Average titre (IU/ngRT) normalised to no drug control

Concentration (μM)

- SAL 4
- SAL 11
- SAL 17
- SAL 12
- SAL 10
- PF74

Figure 62: The effects of modifying the C-terminal cap of SAL-4
(A) A comparison of crystal structures of CPSF6[68]$_{276-290}$ (PDB 4U0A from Price et al. 2014) or SAL-4 in complex with hexameric CA. Peptides shown in red, CA in grey. Red arrows indicate a pocket of CA that the compounds depicted in (B) were designed to reach. **(B)** Chemical structures of SAL-4, 11, 17, 12, or 10. The shaded ovals highlight regions that differ between compounds.. **(C)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74, or the compounds depicted in (B). All infectivity data are representative of at least 2 independent experiments.

phenyl group (SAL-11, 17, 12, and 10, respectively) (figure 62B) either slightly reduced potency or slightly enhanced potency but also increased toxicity (figure 62C). We next tested the effects of adding larger groups to this phenyl (figure 63A). SAL-22 lost almost all antiviral activity, whilst SAL-8 and 16 had similar potency to SAL-4. Interestingly, SAL-7 had increased potency and also exhibited the biphasic phenotype of PF74 (figure 63B).

As well as adding groups to the C-terminal phenyl, we tested the effect of replacing the phenyl group. Replacing it with non-aromatic cyclic groups (SAL-18, 20, and 21) (figure 64A) resulted in a dramatic loss of antiviral activity (figure 64B). Replacing the phenyl group with a 3-pyridyl group (SAL-6) (figure 65A) also resulted in a loss of antiviral activity (figure 65B). We also extended the linker between the phenyl group and the rest of the molecule. Whilst addition of a single CH_2 linker (SAL-5) significantly enhanced the antiviral activity, addition of a CH_2CH_2 linker (SAL-9) reduced the inhibitory effect significantly (figures 65A and B). SAL-5 is a particularly interesting molecule because it inhibits HIV-1 to a similar degree as PF74 at concentrations above $\sim 4\mu M$, but inhibits significantly less well than PF74 at lower concentrations. Of the compounds tested so far, SAL-5 is the most potent single-phase inhibitor and SAL-15 the most post biphasic inhibitor. Surprisingly, all antiviral activity was lost when we combined the two modifications made in SAL-5 and SAL-15 (SAL-23, figures 66A and B).

Table 17 shows the IC_{50}, IC_{90}, and IC_{95} values for the 30 compounds tested. When comparing the potency of these compounds it is important to take into account these three different measurements due to the unusual shaped curves we observed with some compounds.
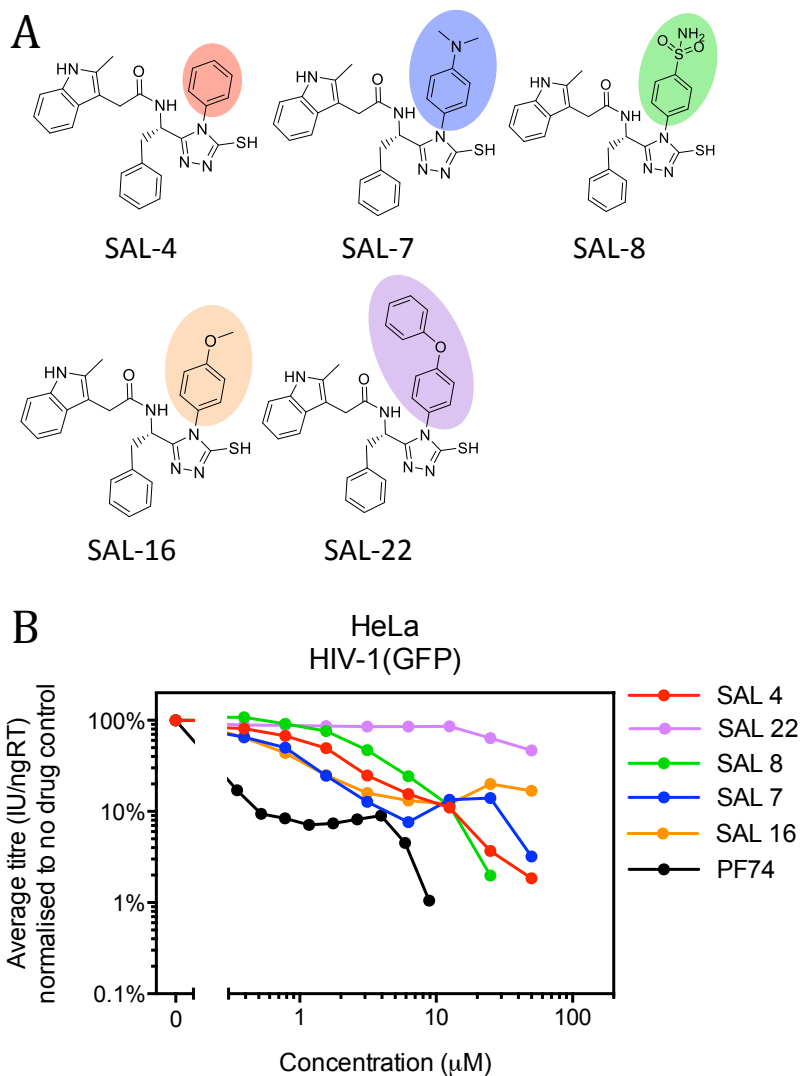
Figure 63: The effects of adding larger groups to the C-terminal cap of SAL-4
(A) Chemical structures of SAL-4, 7, 8, 16, and 22. The shaded ovals highlight regions that differ between compounds. **(B)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74 or the compounds depicted in (A). All data are representative of at least 2 independent experiments.
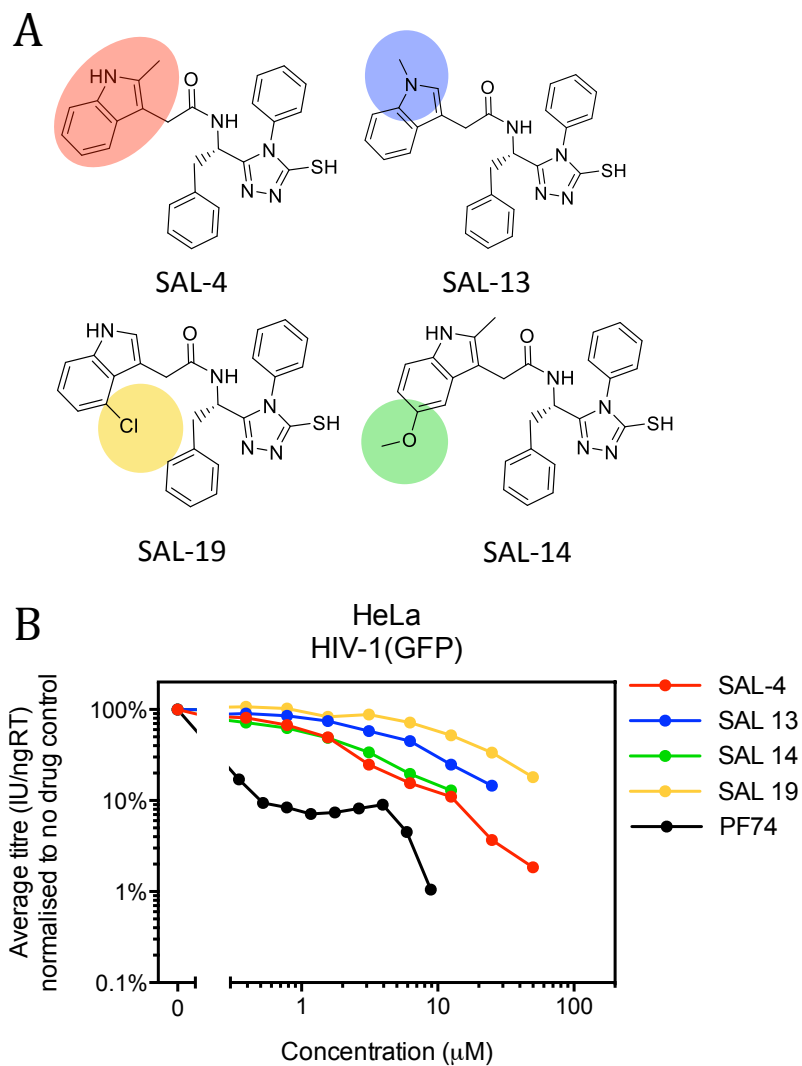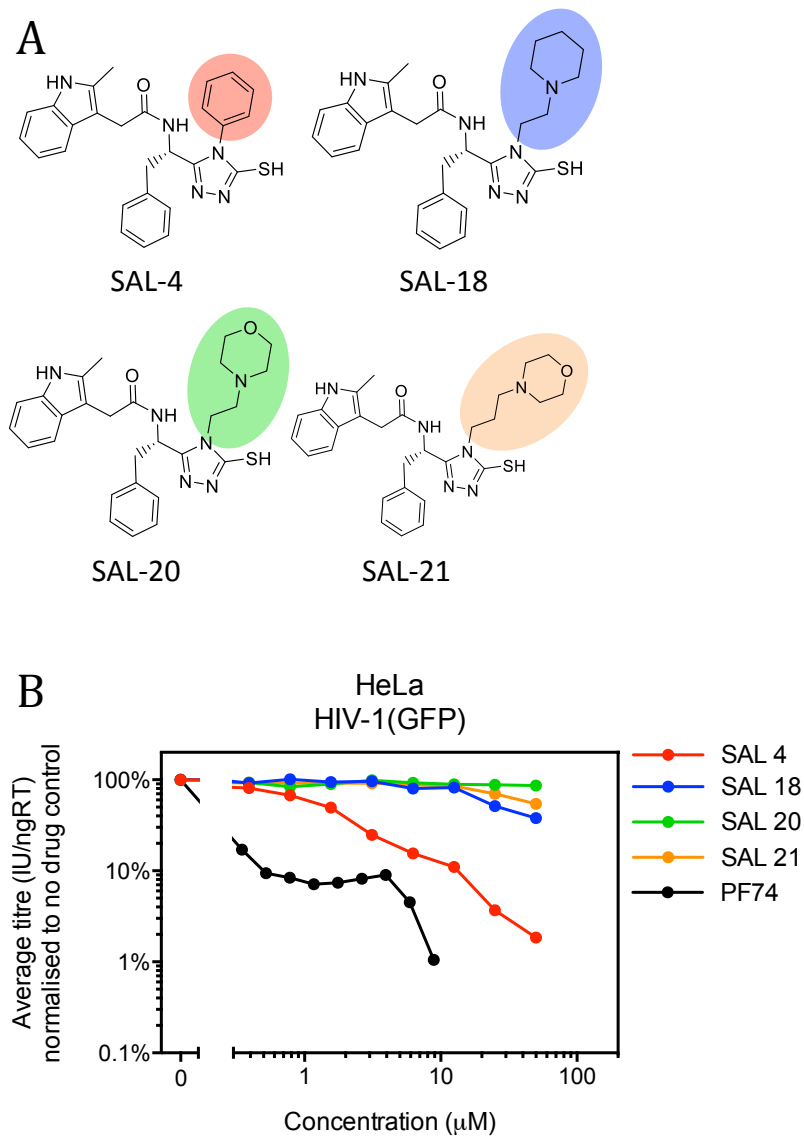
Figure 64: The effects of replacing the C-terminal cap of SAL-4
(A) Chemical structures of SAL-4, 18, 20 and 21. **(B)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74 or the compounds depicted in (A). The shaded ovals highlight regions that differ between compounds. All data are representative of at least 2 independent experiments.
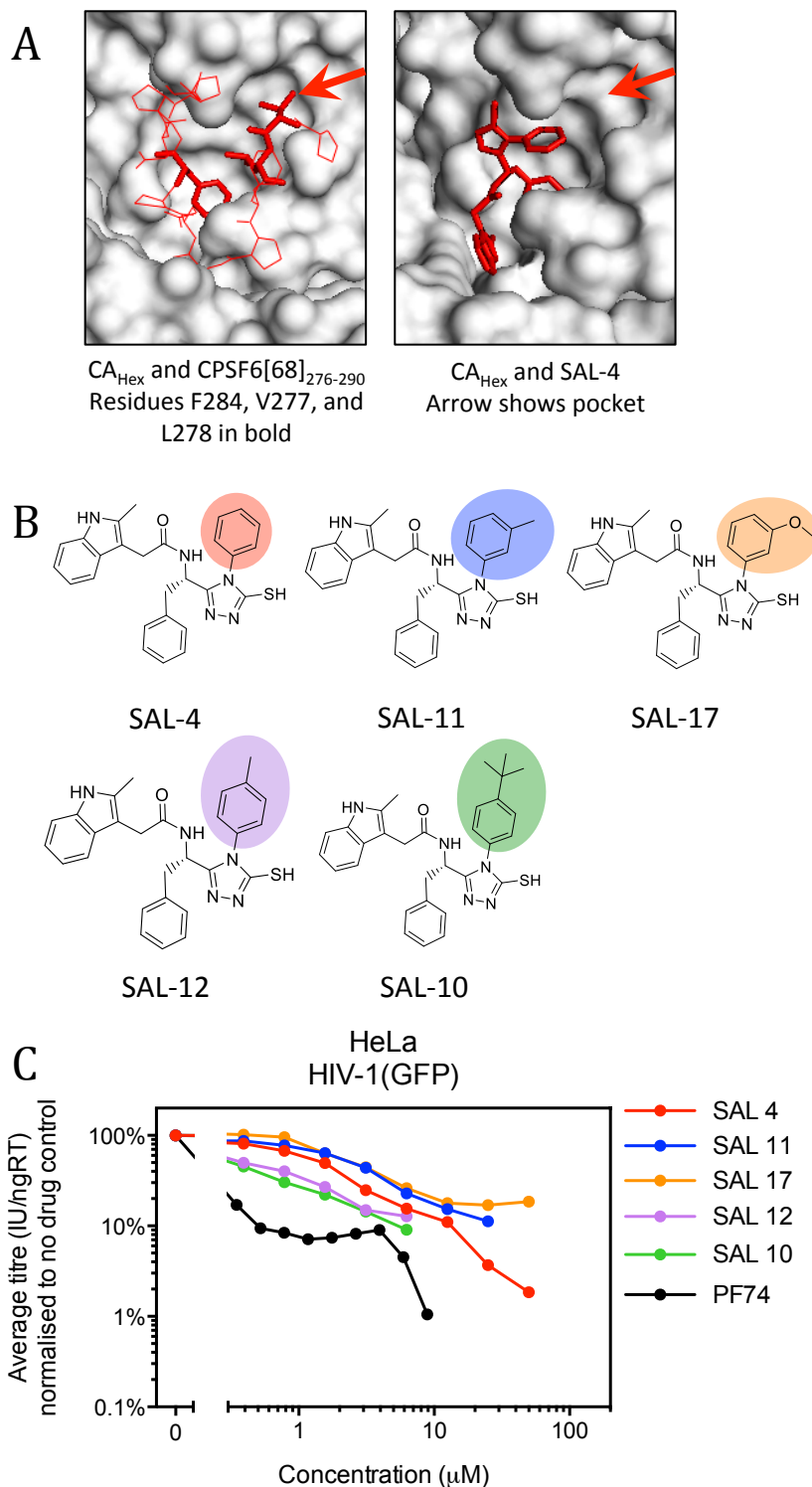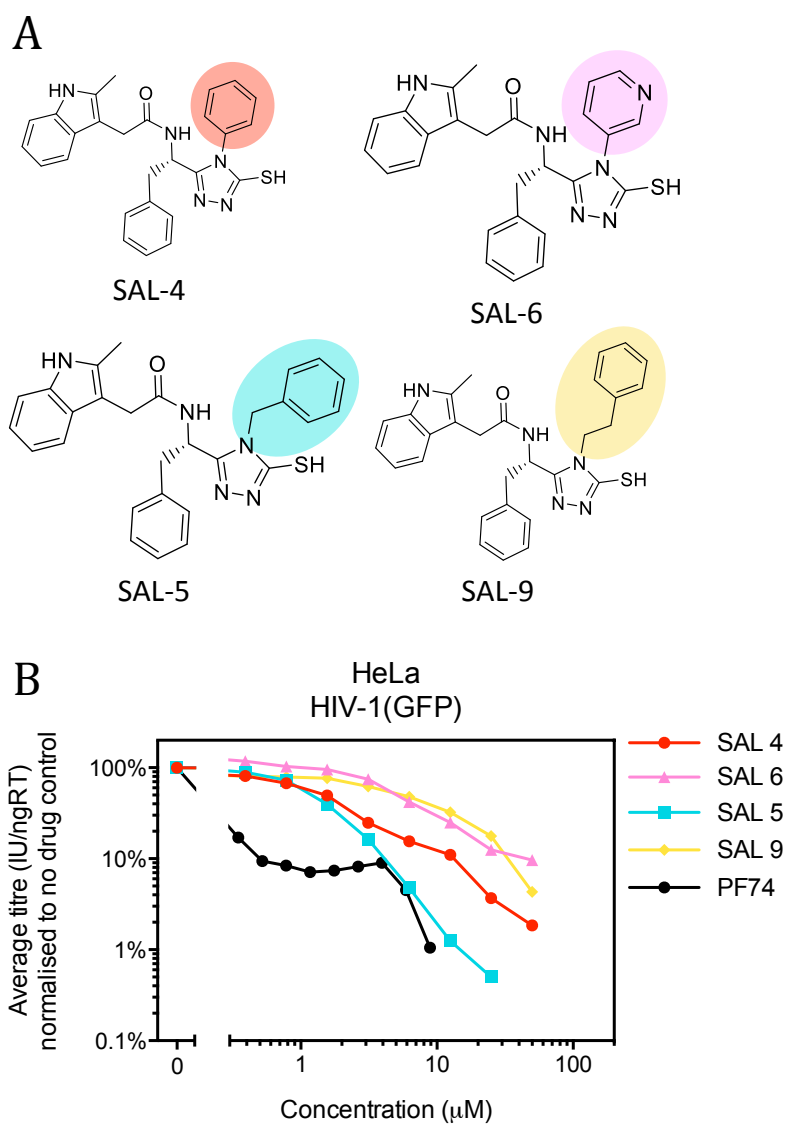
Figure 65: The effects of replacing the C-terminal cap of SAL-4 continued
(A) Chemical structures of SAL-4, 6, 5, and 9. **(B)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74 or the compounds depicted in (A). The shaded ovals highlight regions that differ between compounds. All data are representative of at least 2 independent experiments.
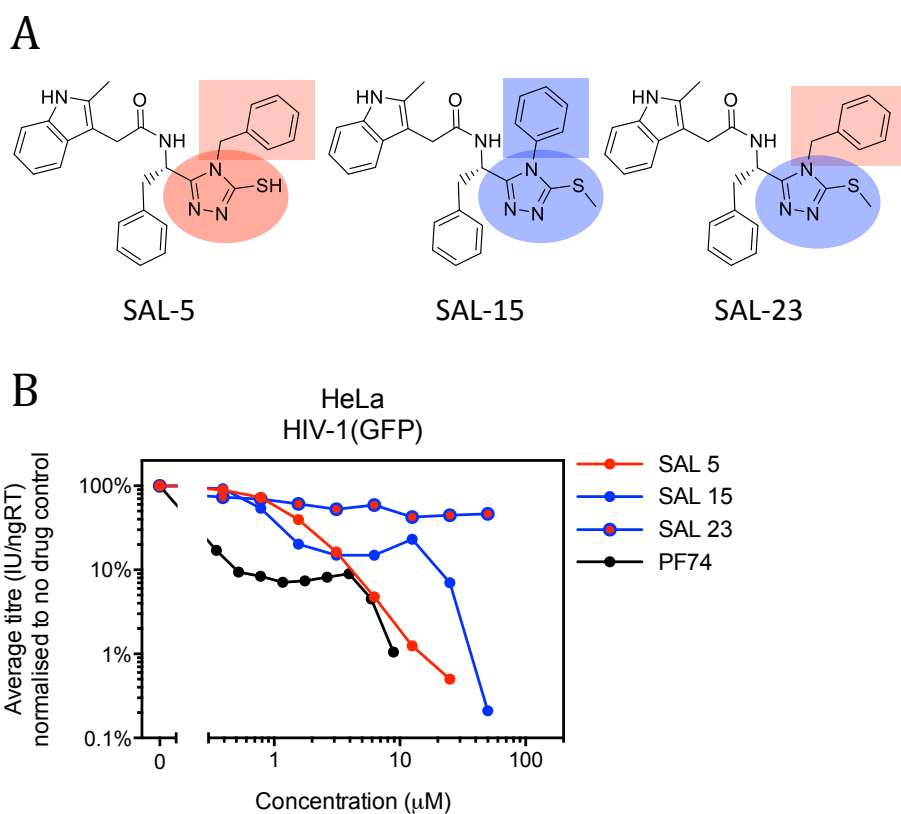
Figure 66: Combining the modifications made to SAL-4 when producing SAL-5 and 15
(A) Chemical structures of SAL-5, 15, and 23. The shaded ovals and rectangles highlight regions that differ between compounds. **(B)** HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74 or the compounds depicted in (A). All data are representative of at least 2 independent experiments.

Drug	IC_{50}	IC_{90}	IC_{95}
SAL 1	57.6		
SAL 2	36.8		
SAL 3	9.6	87.0	
SAL 4	1.5	9.1	24.2
SAL 5	1.3	4.1	6.1
SAL 6	5.4	35.9	
SAL 7	0.7	6.8	40.29
SAL 8	3.0	12.5	19.2
SAL 9	6.0	35.7	50.7
SAL 10	0.3	5.5	13.8
SAL 11	2.5	41.3	
SAL 12	0.4	7.7	23.9
SAL 13	4.7	40.1	133.4
SAL 14	1.4		
SAL 15	1.6	48.7	51.0
SAL 16	1.3		
SAL 17	4.9		
SAL 18	61.8		
SAL 19	27.9		
SAL 20			
SAL 21	58.2		
SAL 22	47.7	72.5	
SAL 23	3.8		
SAL 24	0.4	18.9	24.4
SAL 25	1.4		
SAL 135	2.5		
SAL 143			
SAL 147	3.7		
SAL 151	5.2	45.3	66.9
SAL 157			
PF74	0.2	0.5	5.9
BI-1	6.6	57.4	

Table 17. The $IC_{50, 90}$, and $_{95}$ of PF74, BI-1, and SAL compounds
IC_n = concentration that reduces infectivity by n%. Blank spaces show that this level of inhibition could never be achieved at sub-toxic concentrations of the compound.

5.2.6.2 The ability of SAL compounds to inhibit reverse transcription correlates with their ability to inhibit infection

We previously showed that PF74 only begins to inhibit reverse transcription at doses above ~4μM, at which concentration it inhibits infectivity >11-fold (figure 48D/E). BI-1 never reaches this level of inhibition and so it is possible that it would also be able to inhibit reverse transcription if it could be used at higher doses (which cannot be done due to toxicity and solubility problems). The same could be said for all other CPSF6 peptidomimetics that do not inhibit reverse transcription as well as PF74. If this were true, we would expect to see that the SAL compounds that can inhibit reverse transcription are those that inhibit infectivity most strongly. Any outliers, such as a drug that inhibits infectivity very efficiently without inhibiting reverse transcription at all, would indicate a different mechanism of action. We measured inhibition of infectivity (figure 67A) and late reverse transcripts (figure 67B) for the six most potent compounds identified in this study, alongside PF74 and BI-1. We then compared the ability of each compound to inhibit infection at 20μM with its ability to inhibit reverse transcription at the same concentration. We found a strong positive linear correlation between inhibition of infection and inhibition of reverse transcription (figure 67C). This observation supports the notion that these compounds all inhibit HIV-1 by the same mechanism but, importantly, does not negate the possibility that they work via different mechanisms.

5.2.6.3 SAL-143 causes HIV-1 to become less sensitive to the CPSF6, Nup358, and TNPO3 dependent nuclear entry pathway without inhibiting viral titre

As discussed previously, simply displacing endogenous CPSF6 cannot be the mechanism by which PF74 or BI-1 inhibits the virus because CPSF6 depletion and N74D mutation do not affect titre. We therefore considered it possible to separate the ability of CPSF6 peptidomimetics to displace CPSF6 from their ability to inhibit the virus. In other words, we hypothesised that a compound could be created that displaces endogenous CPSF6 from viral cores without inhibiting the virus. If so, this compound would cause the virus to behave as if
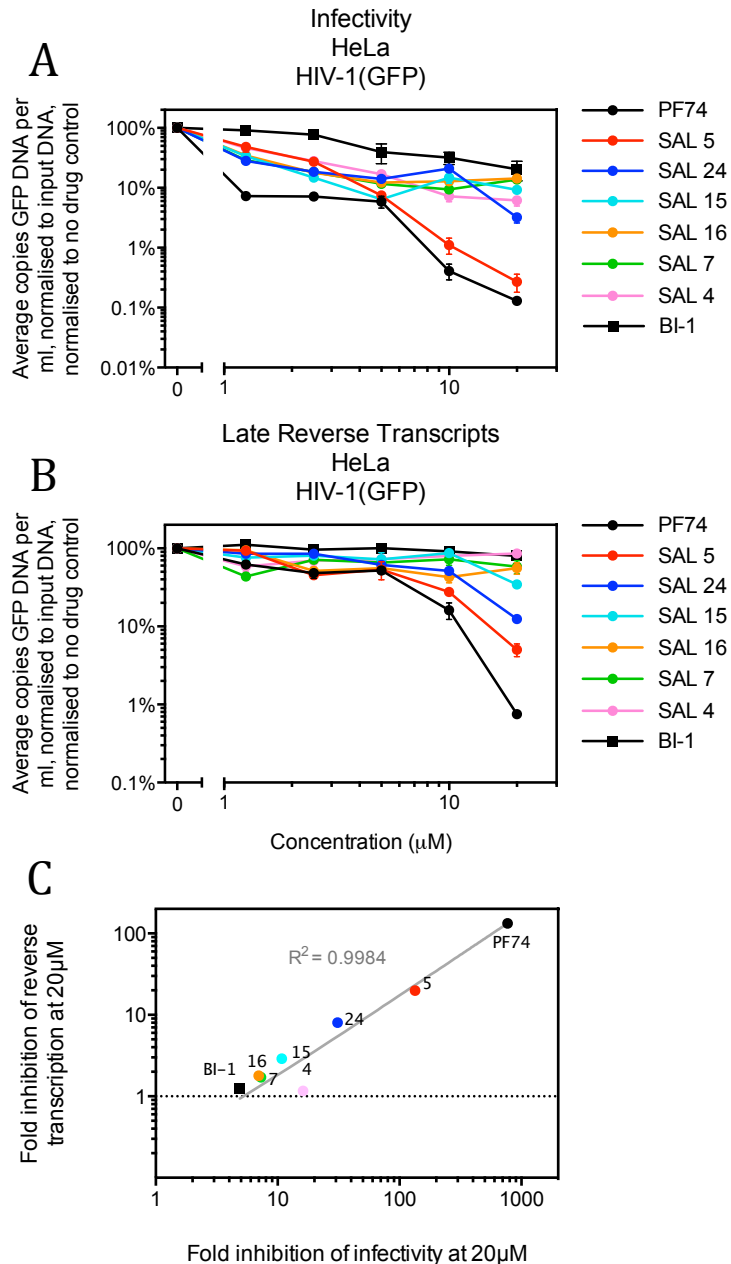
Figure 67. The ability of CPSF6 peptidomimetics to inhibit reverse transcription positively correlates with their ability to inhibit infection
HeLa cells were infected with a titration of HIV-1(GFP) in the presence of a titration of PF74, BI-1, SAL-5, 24, 15, 16, 7, or 4. **(A)** Viral titre was measured at 48hrs p.i. by flow cytometry. **(B)** Late reverse transcripts were measured at 6hrs p.i. by qPCR. **(C)** Data from (A) and (B) shown as fold inhibition of infectivity at 20μM against fold inhibition of reverse transcription at 20μM. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

CPSF6 were depleted or N74 mutated, infecting via a Nup358/TNPO3-independent nuclear entry pathway. An intermediate phenotype may also be possible, in which a compound does have antiviral activity but reduces dependence upon Nup358/TNPO3 at concentrations that do not inhibit the virus.

In the case of PF74 and BI-1, we have shown that insensitivity to Nup358/TNPO3 depletion only occurs at concentrations of the drug that inhibit the virus (figure 49), which simply suggests that PF74/BI-1 inhibit the same stage of HIV-1 infection as Nup358/TNPO3 depletion. We similarly tested the sensitivity of HIV-1 to Nup358/TNPO3 depletion in the presence of our SAL compounds (figures 68 and 69A). For all compounds, we found that a decrease in infectivity correlated with a decrease in sensitivity to Nup358/TNPO3 depletion, as we had seen with PF74 and BI-1. Therefore, when plotting the fold inhibition conferred by a compound at 10μM against the fold inhibition caused by Nup358 or TNPO3 depletion at 10μM of that same compound, we observed a strong negative correlation (figures 69B and C, respectively).

Intriguingly, we found that SAL-143 was an outlier with respect to Nup358 dependence. To confirm his observation, we titrated SAL-143 over a larger range of concentrations. We found that SAL-143 did not inhibit HIV-1 infectivity at all and in fact increased infectivity by up to 1.5-fold. Despite this, 50μM SAL-143 rescued the virus from inhibition by Nup358 and TNPO3 depletion (figure 70A). Nup358 depletion inhibited HIV-1 11-fold in the absence of drug, but only 1.7-fold in the presence of 50μM SAL-143 (a 6.5-fold rescue). Similarly, TNPO3 depletion inhibited HIV-1 28-fold in the absence of drug, but only 7-fold in the presence of 50μM SAL-143 (a 4-fold rescue). We interpreted these observations to mean that SAL-143 allows the virus to enter the nucleus independently of Nup358/TNPO3 by displacing endogenous CPSF6 from viral cores. In order to test this hypothesis, we studied whether SAL-143 could displace HA-hCPSF6[68]-358 from viral cores and thereby rescue the virus from inhibition by this artificial restriction factor. Indeed we found that HA-hCPSF6[68]-358
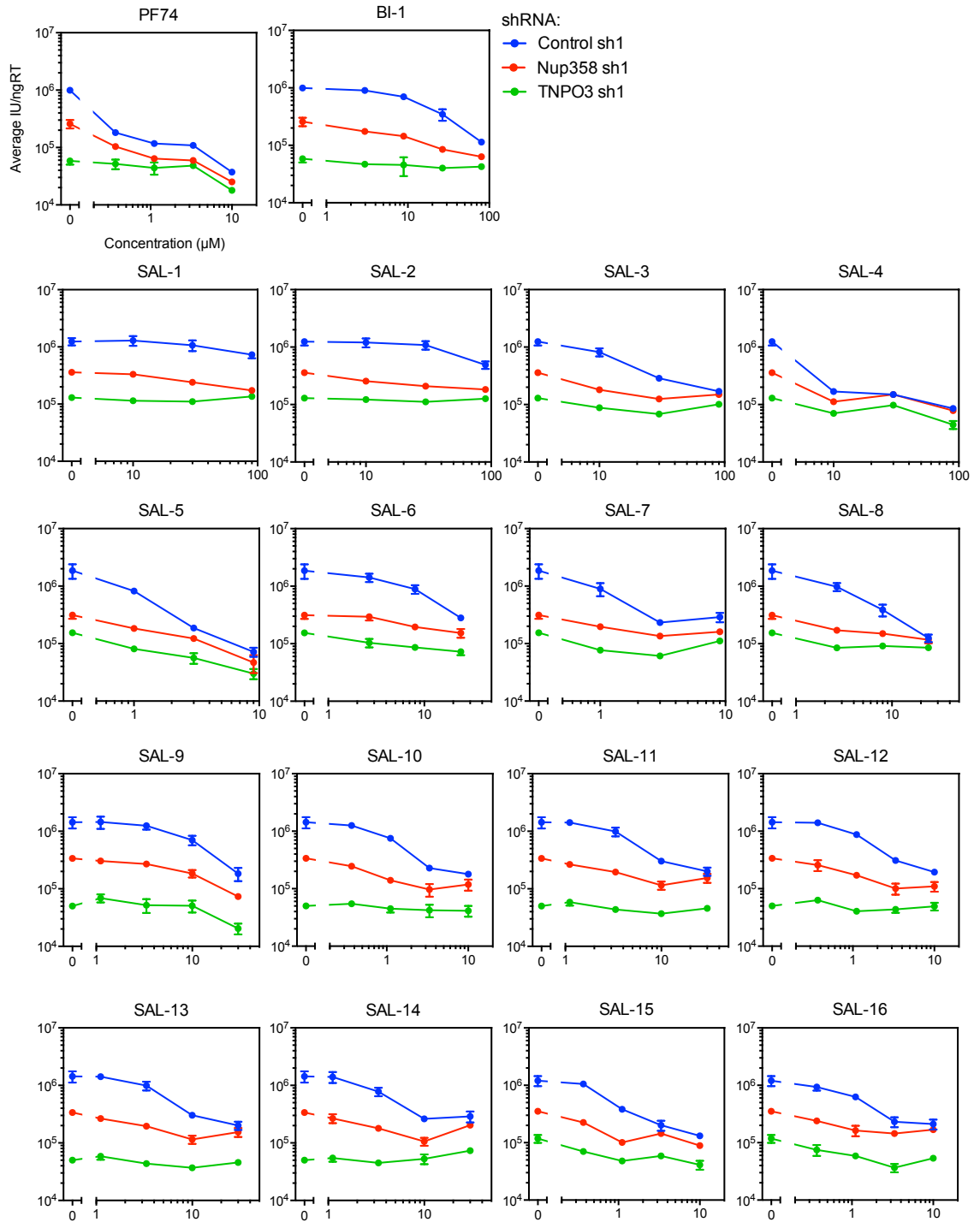
Figure 68. The effect of PF74, BI-1, or SAL-1 to 16 upon the sensitivity of HIV-1 to Nup35 or TNPO3 depletion

HeLa cell clones stably expressing Nup358- or TNPO3-targeting shRNA or a non-targeting control shRNA were infected with a titration of WT HIV-1(GFP) in the presence of a titration of PF74, BI-1, or SAL-1 to 16 and viral titre was measured 48hrs p.i. by flow cytometry. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).
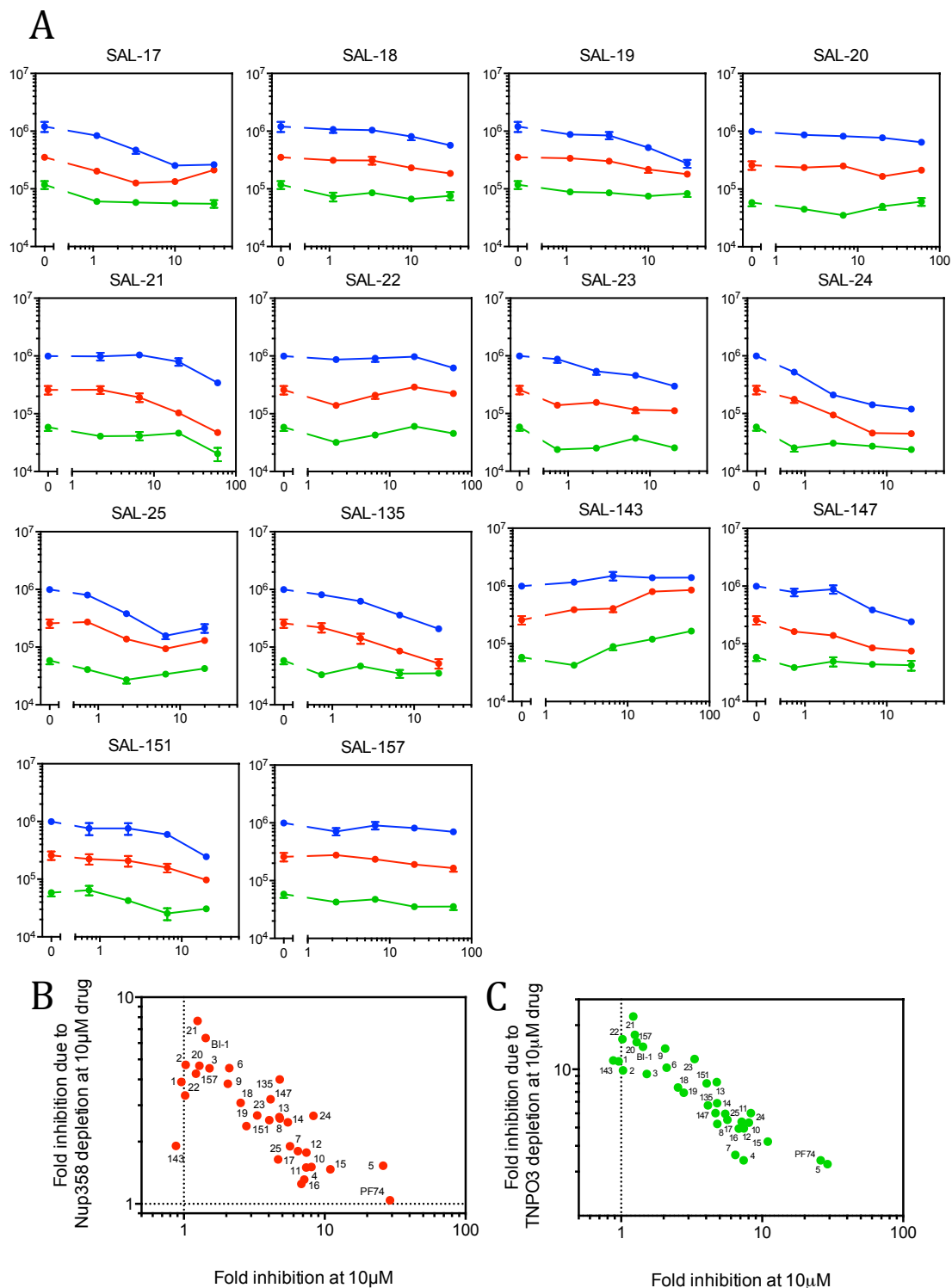
Figure 69. The effect of SAL compounds upon the sensitivity of HIV-1 to Nup358 or TNPO3 depletion

(A) HeLa cell clones stably expressing Nup358- or TNPO3-targeting shRNA or a non-targeting control shRNA were infected with a titration of WT HIV-1(GFP) in the presence of a titration of SAL-17 to 25, 135, 143, 147, 151, or 157 and viral titre was measured 48hrs p.i. by flow cytometry. **(B-C)** show data from (A) as fold inhibition of infection at 10μM against fold inhibition due to (B) Nup358 or (C) TNPO3 depletion at 10μM. Dotted line indicates were fold inhibition is 1. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean).

253
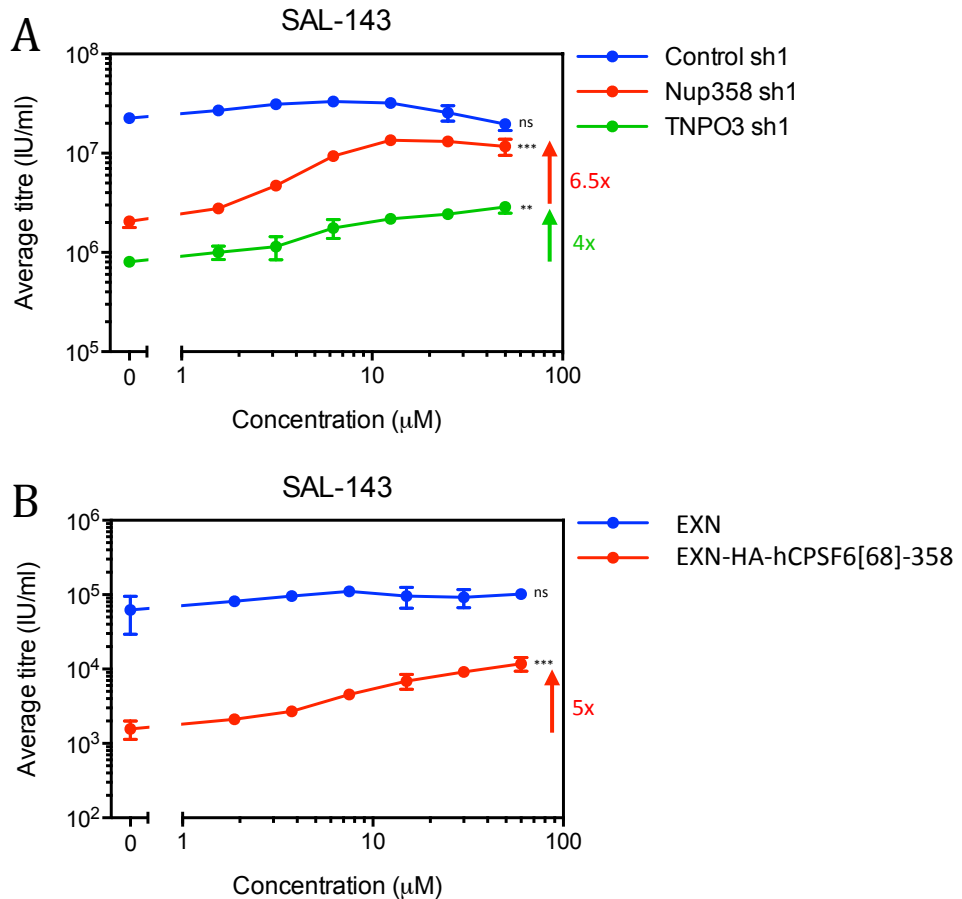
Figure 70. SAL-143 reduces the sensitivity of HIV-1 to Nup358 or TNPO3 depletion and to hCPSF6[68]-358

(A) HeLa cell clones stably expressing Nup358- or TNPO3-targeting shRNA or a non-targeting control shRNA were infected with a titration of WT HIV-1(GFP) in the presence of a titration of SAL-143 and viral titre was measured 48hrs p.i. by flow cytometry. **(B)** HeLa cells stably expressing hCPSF6[68]-358 or empty vector were infected with a titration of WT HIV-1(GFP) in the presence of a titration of SAL-143 and viral titre was measured 48hrs p.i. by flow cytometry. Arrows indicate the fold increase in infectivity caused by 80uM SAL-143. All data are representative of at least 2 independent experiments. Each experiment includes at least 3 biological repeats and error bars show variation between these repeats (bars represent one standard deviation from the mean). ns means p>0.05 (not statistically significant, * means p≤0.05, ** means p≤0.01, *** means p≤0.001, as measured by Student's *t*-test. Statistics show whether the viral titre at 60uM SAL-143 is significantly different from the viral titre in the absence of drug (in the same cell line).

restricted HIV-1 40-fold in the absence of drug but only 8.5-fold in the presence of 50μM SAL-143 (a 5-fold rescue) (figure 70B).

5.2.6.4 *SAL-5 inhibits HIV-1 at a lower CA occupancy than other CPSF6 peptidomimetics*

We previously demonstrated that PF74 and BI-1 inhibit HIV-1 at a higher CA occupancy than HA-hCPSF6[68]-358, which suggests that they may inhibit via a different mechanism of action (figure 52). We hypothesised that compounds that mimic CPSF6 more closely may inhibit the virus at a similar CA occupancy. To test whether our eight most potent compounds also inhibited at higher occupancies than HA-hCPSF6[68]-358, we titrated them onto HeLa cells expressing this factor. With the exception of SAL-143 and SAL-5, we found that all compounds were able to rescue infectivity from HA-CPSF6[68]-358 in a similar manner to PF74 and BI-1 (figure 71). We had previously noticed that the inhibition curve of SAL-5 differed from other compounds with similar potencies because it was not biphasic (figures 65B). Here, we found that SAL-5 did not rescue the virus from HA-hCPSF6[68]-358. We interpret this to mean that SAL-5 inhibits at the same (or even lower) CA occupancy as HA-hCPSF6[68]-358.
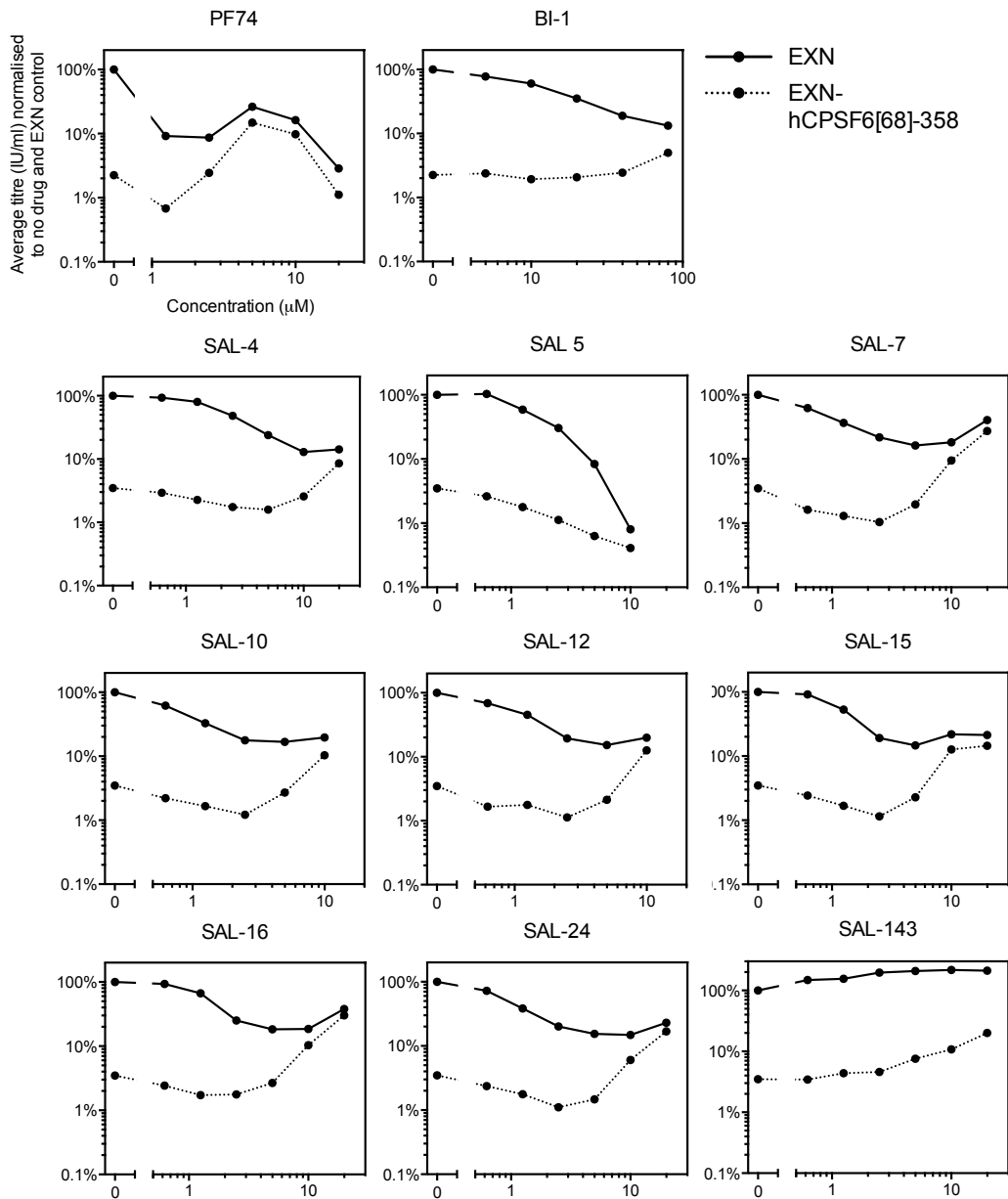
Figure 71. SAL-5 inhibits HIV-1 at an occupancy the same as or lower than that of hCPSF6[68]-358

HeLa cells stably expressing hCPSF6[68]-358 or empty vector were infected with a titration of WT HIV-1(GFP) in the presence of a titration of PF74, BI-1, SAL-4, 5, 7, 10, 12, 15, 16, 24, or SAL-143 and viral titre was measured 48hrs p.i. by flow cytometry. All data are representative of at least 2 independent experiments.

5.3 Discussion

5.3.1 HA-hCPSF6[68]-358, PF74, and BI-1 inhibit the normal nuclear entry pathway of HIV-1 through reversible, CypA-dependent mechanisms

As discussed in section 3.3.3, several studies have examined the inhibition of HIV-1 by cytoplasmic antiviral forms of CPSF6 (247, 259, 342-345). The mechanism of inhibition is still in dispute, not least because different constructs inhibit at different stages of infection. It has been suggested that constructs inhibiting reverse transcription inhibit through a different mechanism than constructs inhibiting nuclear entry (344), but this finding has not yet been reproduced. Instead, the consensus is that antiviral forms of CPSF6 inhibit HIV-1 by stabilising cores and preventing them from uncoating (90, 247, 345). However, this does not explain why some constructs inhibit reverse transcription and other do not. One possibility is that constructs differ in the degree of stability they confer (due to expression levels, CA affinity, and/or CA occupancy) and a certain degree of stability must be reached in order to inhibit reverse transcription. Another possibility is that constructs differ in the level of CA occupancy they achieve and that, at a certain occupancy, CPSF6 obstructs pores within the HIV-1 core, making it impervious to dNTPs and thus inhibit reverse transcription.

These hypotheses not only apply to different antiviral forms of CPSF6 but also to CPSF6 peptidomimetics. These small-molecule compounds are particularly useful tools because, unlike proteins, they can easily be titrated and so we have been able to test the hypothesis that the stage of inhibition is concentration-dependent. Blair, et al. (411) reported that PF74 inhibited HIV-1 prior to reverse transcription but, upon titrating PF74, we found that it has an unusual biphasic dose response curve, which plateaus between approximately 0.5 and 4μM, and that reverse transcription was only inhibited above ~4μM. This was consistent with the hypothesis that molecules interacting with the CPSF6 binding pocket of CA can have different effects at different concentrations. In agreement with Lamorte, et al. (327), we found that BI-1 cannot inhibit reverse

transcription at any non-toxic dose. However, whilst Lamorte, et al. (327) interpret the different inhibition profiles of PF74 and BI-1 to mean that they inhibit by fundamentally different mechanisms, we hypothesise that BI-1 inhibits HIV-1 by the same mechanism as PF74 and simply does not reach the second phase of inhibition at sub-toxic concentrations (due to its lower affinity for CA and/or its poor solubility). In order to better understand the inhibition profiles of these two molecules we further probed their mechanism of action.

Thus far, the biochemical assays used to study the effects of PF74 and BI-1/2 upon core stability have produced inconsistent results (87, 90, 113, 327). As discussed in section 3.3.2, these studies have many caveats, including the inability to differentiate between virions that would go on to be infectious and those that would not. We therefore used a drug washout assay to study the effects of PF74 or BI-1 upon core stability in the context of an infection and used infectivity as a read-out. Through this assay we demonstrated that both PF74 and BI-1-mediated inhibition of HIV-1 infection and/or reverse transcription is reversible. We presume that uncoating is an irreversible process because the process of core assembly and maturation in producer cells requires such complex machinery (discussed in section 1.2.11). These data strongly argue that neither PF74 nor BI-1 induces premature uncoating of HIV-1. Similarly, we demonstrated that HA-hCPSF6[68]-358-mediated inhibition of reverse transcription is reversible by displacing this artificial restriction factor from viral cores using concentrations of PF74 or BI-1 that do not inhibit reverse transcription. This strongly argues against the notion C-terminal truncations of CPSF6 inhibit reverse transcription by prematurely uncoating cores. Importantly, the fact that these molecules do not uncoat cores does not necessarily mean that they stabilise cores – they may not affect stability at all. However, the hypothesis that CPSF6, and therefore its peptidomimetics, stabilise cores is strongly supported by the recent finding that endogenous CPSF6 'cloaks' HIV-1 reverse transcripts and prevents them from being exposed to PRRs in MDMs (92) (discussed in section 3.3.7).

The finding that concentrations of PF74/BI-1 that do not inhibit reverse transcription were able to displace (at 6 hours post-infection) a concentration of HA-hCPSF6[68]-358 that does inhibit reverse transcription suggested that PF74/BI-1 and HA-hCPSF6[68]-358 inhibit reverse transcription at different CA occupancies. We investigated this phenotype further and found that PF74/BI-1 was able to rescue both infectivity and reverse transcription from inhibition by HA-hCPSF6[68]-358 when added at the same time as infection. We interpret this to mean that HA-hCPSF6[68]-358 inhibits HIV-1 at a lower CA occupancy than PF74/BI-1. In other words, fewer molecules of HA-hCPSF6[68]-358 than PF74/BI-1 must be bound to a viral core in order to confer the same level of inhibition. This either means that HA-hCPSF6[68]-358 inhibits through different mechanisms from PF74/BI-1 or that they inhibit through the same mechanism but with different efficiencies. The latter could be possible if, for example, all three inhibitors stabilise cores and prevent uncoating but HA-hCPSF6[68]-358 stabilises the core more effectively than PF74/BI-1, perhaps because it makes more extensive contacts with CA.

We, and others, have shown that PF74 and BI-1 can displace CPSF6 from CA (figures 51 and 52) (83, 413) and that the inhibitory activity of these small-molecules cannot simply be a result of their ability to displace CPSF6 (figure 20)(342). We therefore considered the possibility that PF74/BI-1 might displace endogenous CPSF6 at sub-inhibitory concentrations. Having previously shown that the consequence of disrupting CA-CPSF6 interaction is insensitivity to Nup358 or TNPO3 depletion, we tested whether PF74 and BI-1 were able to confer insensitivity to Nup358/TNPO3 at sub-inhibitory concentrations. However, we only observed insensitivity to Nup358/TNPO3-depletion at inhibitory concentrations PF74/BI-1. This suggests that PF74 and BI-1 inhibit the virus at concentrations equal to or below concentrations that displace endogenous CPSF6. The fact that PF74/BI-1 and Nup358/TNPO3-depletion are not additive in their ability to inhibit the virus also supports the hypothesis that they inhibit the same stage/pathway of infection.

Whatever the mechanism of action of HA-hCPSF6[68]-358, PF74, or BI-1, it appears to be dependent upon CA-CypA interaction. Shi, et al. (113) previously reported that CsA or depletion of CypA antagonised PF74 and Hilditch (unpublished observations) similarly found that SmBz (a non-immunosuppressive CsA analog) reduced the ability of HA-hCPSF6[68]-358 to inhibit HIV-1. We previously hypothesised that CypA stabilises cores and that dissociation of CypA from cores was dependent upon CPSF6 trafficking cores to the NPC, where Nup358 competes off CypA. If so, cores inhibited by HA-hCPSF6[68]-358/PF74 would not traffic to the nucleus and so CypA-mediated stabilisation would be permanent and inhibitory. It follows that the inhibitory effects of HA-hCPSF6[68]-358/PF74 and CypA would be additive and, conversely, that CsA would appear to antagonise HA-hCPSF6[68]-358/PF74. Alternatively, these data could be interpreted to mean that CA-CypA interaction causes conformational changes that encourage CA to interact with HA-hCPSF6[68]-358/PF74.

In agreement with these previous findings, we found that 10μM SmBz antagonised low concentrations of PF74 in HeLa cells. However, we found that SmBz was synergistic with high concentrations of PF74. Furthermore, SmBz never antagonised PF74 in TE671s. Instead SmBz and PF74 were either additive or synergistic, in a PF74 concentration-dependent manner. In both cell types, the presence of SmBz caused PF74 to lose its characteristic biphasic inhibition curve, which could be interpreted to mean that the plateau in the PF74 inhibition curve is somehow due to CypA. Importantly, SmBz and PF74 were additive in their ability to inhibit reverse transcription in both cell types. When testing the effect of SmBz upon BI-1, we observed the same phenotype as with low doses of PF74. This complex relationship between SmBz and PF74/BI-1 is difficult to interpret because the mechanism of action of neither CypA, CPSF6, nor their inhibitors has been determined and because the effects of CypA/SmBz are cell type-dependent. However, given the similar roles of CypA and CPSF6 in early HIV infection, it is unsurprising that they influence each other. These observations suggest that the cell-type specificity of CypA, and also CPSF6, warrants further investigation.

Based on the observations made in this study and others, we have proposed a model in which endogenous CPSF6 binds to cores and stabilises them, preventing uncoating. CPSF6 then traffics cores to the NPC where it dissociates, allowing cores to uncoat at the NPC. We propose that CPSF6 C-terminal truncations, PF74, and BI-1 mimic CPSF6 in its ability to bind CA and to stabilise cores but not in its ability to traffic to the NPC or dissociate from cores. Therefore, cores are permanently stabilised in the cytoplasm and are unable to enter the nucleus. We hypothesise that HA-hCPSF6[68]-358 and high concentrations of PF74 inhibit reverse transcription either by stabilising cores such that reverse transcripts cannot grow or by obstructing pores in viral cores such that dNTPs cannot access the reverse transcribing genome. We propose that BI-1 and low concentrations of PF74 do not inhibit reverse transcription because they have not achieved the CA occupancy required to do so. We suggest that HA-hCPSF6[68]-358 inhibits at a lower CA occupancy than PF74/BI-1 due to its size, which means that it makes more extensive contacts with CA than PF74/BI-1 and occludes entry of dNTPs more efficiently than PF74/BI-1.

5.3.2 Molecules interacting with the CPSF6-binding pocket of CA can have entirely different effects upon the virus

Having examined the mechanisms by which HA-hCPSF6[68]-358, PF74, and BI-1 inhibit HIV-1 infection, we sought to explore the effect of other ligands interacting with this CA pocket. We therefore designed a series of small molecules based on the structure of PF74 and compared their phenotypes with PF74/BI-1. Although this was primarily an exploratory study, we did design molecules rationally. For example, we tried to increase the potency of the compound by adding small functional groups that we predicted would pick up additional interactions and increase its affinity for CA. We also tried to make compounds that mimicked CPSF6 more closely than PF74 does by extending the C-terminus of the molecule into pockets reached by CPSF6 but not PF74/BI-1. We also cyclised the core scaffold such the compound was less flexible and was permanently orientated in its bound conformation. We were also particularly

interested in the concept of creating a compound that caused uncoating of the virus, exposing viral DNA to innate PRRs and triggering an antiviral response in primary MDMs. We therefore also designed compounds that were significantly larger than PF74, with the aim of disrupting CA-CA interactions and causing disassembly.

The core scaffold of PF74 is an amide group that is found primarily in the *trans* conformation in solution but adopts the *cis* conformation in complex with CA (340). We therefore created a series of compounds in which this amide was cyclised, fixing it in the *cis* conformation. A crystal structure of SAL-4 in complex with hexameric CA confirmed that it adopts an almost identical conformation as PF74. Whilst some of these compounds had a similar potency to PF74 at high doses (>5μM), for example SAL-5, PF74 was the most potent compound by far at lower concentrations. Cyclising this bond reduces the flexibility of the compound, which could be the cause of the reduced antiviral activity. Reduced flexibility may also cause compounds to be more likely to disrupt this pocket of CA, which would be beneficial if aiming to create a compound that disassembles cores.

The majority of compounds screened in this study showed similar phenotypes to PF74 and/or BI-1. We observed a strong positive correlation between the ability of compounds to inhibit reverse transcription and their ability to inhibit infectivity. Like PF74, the degree of inhibition of reverse transcription could never account for the degree of inhibition of infection, supporting the hypothesis that reverse transcription is inhibited at a higher CA occupancy than infectivity. We found that the ability of almost all compounds to reduce sensitivity to Nup358/TNPO3-depletion correlated with their ability to inhibit infection (with the exception of SAL-143, discussed below). Similarly, almost all compounds tested were able to rescue HIV-1 from HA-hCPSF6[68]-358-mediated inhibition in a concentration dependent manner, showing that they inhibit the virus at a lower occupancy than this artificial restriction factor (with the exception of SAL-5, discussed below).

It remains unclear why PF74, but not BI-1, has a biphasic inhibition curve. We, and others, have proposed that BI-1 would exhibit such a phenotype if it could be used at a high enough concentration to reach the same level of inhibition as PF74. Whilst this may be true for BI-1, we have identified several compounds that show that this is not true for all antivirals interacting with this pocket. For example, SAL-4 has a normal inhibition curve but we have created several compounds of similar potencies to SAL-4 (when potency is measured taking into account IC_{50}, IC_{90}, and IC_{95}) that exhibit a biphasic inhibition curve. These include SAL-15 and 24, which differ from SAL-4 through the functional groups attached to the cyclised core, and SAL-7 and 16, which differ from SAL-4 through the functional groups attached to the C-terminal phenyl cap. Similarly, SAL-5 reaches the same potency as PF74 at $\sim 4\mu M$ and yet does not exhibit a biphasic curve. Importantly, both compounds with one and two-phase inhibition curves were able to inhibit reverse transcription. It would be of great interest to compare the crystal structures of compounds with these two different inhibition phenotypes.

SAL-5 is of particular interest because, unlike all other compounds tested, it cannot rescue HIV-1 infectivity from HA-hCPSF6[68]-358. This suggests that SAL-5 inhibits at a similar, or even lower, occupancy than HA-hCPSF6[68]-358 and thus a lower occupancy than the other compounds tested. SAL-5 differs from SAL-4 due to an extra CH_2 linker between the C-terminal phenyl cap and the rest of the molecule. Based upon the structure of SAL-4 in complex with CA (figure 58B), we predict that this extra linker allows the phenyl group of SAL-5 to extend further into a channel of CA that does not interact with PF74/BI-1/SAL-4 but interacts extensively with CPSF6[68]$_{276-290}$. In this way, SAL-5 may more closely mimic CPSF6 than other compounds.

In light of the discovery that CPSF6[68]$_{276-290}$, PF74, and SAL-4 all interact with two adjacent CA monomers within a hexamer, we were particularly interested in exploring modifications of the region of PF74 or SAL-4 that interacts with this second monomer, the indole group. Thus far, all compounds in which we replaced the indole group or added functional groups to the indole resulted in a

significant loss of antiviral activity. SAL-143, in which we replaced the indole group with an oxindole, enhanced infectivity by up to 2-fold. Further investigation into the effects of SAL-143 upon the nuclear entry pathway of HIV-1 revealed that it caused HIV-1 to become 6.5-fold less sensitive to Nup358 depletion, 4-fold less sensitive to TNPO3 depletion, and 5-fold less sensitive to HA-hCPSF6[68]-358-mediated restriction. We interpret these observations to mean that SAL-143 induces core uncoating and allows the virus to infect via the efficient CPSF6/Nup358/TNPO3-independent nuclear entry pathway, like N74D. If so, we would predict that this compound would cause WT HIV-1 (or at least a proportion of WT HIV-1) to be detectable in MDMs, thereby inducing an antiviral response. Although the phenotype of SAL-143 is partial, it is an excellent proof of principle, which encourages us to screen more compounds based upon the structure of SAL-143. Furthermore, only a small proportion of virions may need to be detected by the innate immune system in order to mount an efficient antiviral response. A compound that causes HIV-1 to trigger an innate immune response *in vivo* could theoretically function as a prophylactic by inducing a local antiviral response to viruses during transmission, preventing their replication and spread.

6   **Chapter 6. Future work**

Our model of a HIV-1 nuclear entry could be further tested in several ways. For example, the immunofluorescence assays used by McDonald *et al.* (65) or the superresolution microscopy assays used by Lelek *et al.* (101) to study RTC composition during early HIV-1 infection could be used to test whether NES-CPSF6[68] and CPSF6 peptidomimetics prevent core uncoating, as has been suggested in this study and others. These assays could also be used to test the effect of cofactor depletion and/or CA mutation upon the composition and location of the virus. If CA mutants whose nuclear entry pathway has been well characterised in HeLa cells – such as N74D and R132K/L136M – were found to be phenotypically different from WT HIV-1 in these assays, their phenotypes could also be studied in CD4$^+$ T cells and MDMs as a way of examining the nuclear entry pathway in these cell types. Whilst there is a significant amount of evidence that Nup358, TNPO3, and Nup153 are cofactors of HIV-1 nuclear import in HeLa cells, their involvement in nuclear entry in primary cells has not been well studied, primarily because these cells are very difficult to manipulate through protein depletion or expression. Microscopy techniques could therefore be an excellent way of comparing the nuclear entry pathway of HIV-1 in cell lines and primary cells.

We, and others, have demonstrated that the nuclear entry pathway taken by HIV-1 is reflected in its integration site characteristics. Therefore, characterisation of WT and mutant HIV-1 integration sites in primary cells could help to determine whether the virus takes similar nuclear entry pathways in these cells as in HeLa cells. It would also be important to test the location of these proviruses within the nucleus of primary cells using the HIV-1 LTR DNA FISH assay designed by Lusic *et al.* (unpublished observations) because these cell types likely differ from tumorigenic HeLa cells in their genomic organisation.

Our study of B*27-associated CTL escape mutants has suggested that the mutations R132K and S41A alter the interaction of CA with CypA, Nup358, and

CPSF6, and ITC could now be used to measure these interactions directly. If ITC data proved consistent with our observations, then the fact that S41A/R132K/L136M does not integrate into the expected integration sites would suggest that CA-dependent factors other than CypA, CPSF6, Nup358, and TNPO3 are influencing integration site targeting. We have developed a SureSelect[XT2] assay for integration site sequencing and this could now be used to screen the effects of a series of other nucleoporins and karyopherins upon integration site selection of HIV-1.

Having found that at least one mutation other than S41A can compensate for the defect of R132K *in vivo*, the candidate mutations we identified could now be tested. Examining the effects of this alternative compensatory mutation upon the nuclear entry pathway of HIV-1 could be very informative, particularly if it rescues R132K through a different mechanism than S41A.

Whilst this study has examined the molecular interactions that HIV-1 makes with nuclear import cofactors, others have focused upon the consequences of these interactions with respect to innate immune detection in MDMs. In order to understand the mechanism by which CypA and CPSF6, and potentially other cofactors, determine innate immune detection of the virus, it would be extremely useful to be able to study both aspects in the same cells. Unfortunately, HeLa cells cannot be used to study innate detection because they have defective innate signaling pathways (Tan CP., unpublished observations) and MDMs cannot easily be used to study the molecular details of the nuclear entry pathway because they are so difficult to manipulate. Existing cell lines could be screened for the ability to detect, for example, N74D and P90A. Alternatively, sequencing and protein expression studies could be used to determine the point/s of the cGAS signaling pathway at which HeLa cells are defective so that it could be reconstituted to detect P90A. Of course, the innate immune sensor of N74D in MDMs would need to be identified before the same was possible for this mutant. If this were achieved, CA mutants and cofactor depletions known to influence HIV-1 nuclear entry could easily be tested for their ability to trigger an innate immune response.

Such a cell line could be used to test whether SAL-143, and other CPSF6 peptidomimetics, exposes WT HIV-1 to the innate immune system as we predict it would. Whilst MDMs can also be used to test these compounds, screening of large numbers of compounds would be much more efficient and reproducible in a reporter cell line. Regardless of whether antiviral CPSF6 peptidomimetics cause the virus to trigger an IFN response or not, we hypothesise that escape mutants of these compounds would trigger an IFN response. To test this, viral outgrowth assays could be used to identify escape mutants, which could subsequently be screened for their ability to induce an IFN response. The notion of creating an antiviral compound whose escape mutants are unable to replicate because they induce an antiviral response themselves is a particularly exciting prospect.

7 References

1. **Bishop JM.** 1978. Retroviruses. Annu. Rev. Biochem. **47**:35-88
2. **Baltimore D.** 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature **226**(5252):1209-1211
3. **Temin HM, Mizutani S.** 1992. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. 1970. Biotechnology **24**:51-56
4. **The International Committee on Taxonomy of Viruses.** 2012. Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego.
5. **Vogt VM, Eisenman R.** 1973. Identification of a large polypeptide precursor of avian oncornavirus proteins. Proc. Natl. Acad. Sci. USA **70**(6):1734-1738
6. **Vogt VM.** 1997. Retroviral virions and genomes, p. 27-70. *In* Coffin JM, Hughes SH, Varmus HE (ed.), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).
7. **Weiss RA**. 2006. The discovery of endogenous retroviruses. Retrovirology **3**(67) [Online] doi: 10.1186/1742-4690-3-67
8. **Tang H, Kuhen KL, Wong-Staal F.** 1999. Lentivirus replication and regulation. Annu. Rev. Genet. **33**:133-170
9. **Yamashita M, Emerman M.** 2006. Retroviral infection of non-dividing cells: old and new perspectives. Virology **344**(1):88-93
10. **CDC**. 1981. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men - New York City and California. **30**(25):305-308. MMWR Weekly
11. **Hymes KB, Cheung T, Greene JB, Prose NS, Marcus A, Ballard H, William DC, Laubenstein LJ.** 1981. Kaposi's sarcoma in homosexual men - a report of eight cases. Lancet **2**(8247):598-600
12. **Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, Saxon A.** 1981. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. New Engl. J. Med. **305**(24):1425-1431
13. **Waterson AP.** 1983. Acquired immune deficiency syndrome. Br. Med. J. (Clin. Res. Ed.) **286**(6367):743-746
14. **Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220**(4599):868-871
15. **Popovic M, Sarngadharan MG, Read E, Gallo RC.** 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science **224**(4648):497-500
16. **Gallo RC, Montagnier L.** 2003. The discovery of HIV as the cause of AIDS. New Engl. J. Med. **349**(24):2283-2285

17. **Coffin J, Haase A, Levy JA, Montagnier L, Oroszlan S, Teich N, Temin H, Toyoshima K, Varmus H, Vogt P, et al.** 1986. What to call the AIDS virus? Nature **321**(6065):10

18. **Sonigo P, Alizon M, Staskus K, Klatzmann D, Cole S, Danos O, Retzel E, Tiollais P, Haase A, Wain-Hobson S.** 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell **42**(1):369-382

19. **Gonda MA, Wong-Staal F, Gallo RC, Clements JE, Narayan O, Gilden RV.** 1985. Sequence homology and morphologic similarity of HTLV-III and visna virus, a pathogenic lentivirus. Science **227**(4683):173-177

20. **Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey MA, Santos-Ferreira MO, Laurent AG, Dauguet C, Katlama C, Rouzioux C, et al.** 1986. Isolation of a new human retrovirus from West African patients with AIDS. Science **233**(4761):343-346

21. **Masur H, Michelis MA, Greene JB, Onorato I, Stouwe RA, Holzman RS, Wormser G, Brettman L, Lange M, Murray HW, Cunningham-Rundles S.** 1981. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. New Engl. J. Med. **305**(24):1431-1438

22. **CDC**. 1982. Epidemiologic notes and reports Pneumocystis carinii pneumonia among persons with hemophilia A. **31**(27):365-367. MMWR Weekly

23. **CDC**. 1982. Unexplained immunodeficiency and opportunistic infections in infants - New York, New Jersey, California. **31**(49):665-667. MMWR Weekly

24. **Morison L.** 2001. The global epidemiology of HIV/AIDS. Br. Med. Bull. **58**:7-18

25. **Shaw GM, Hunter E**. 2012. HIV transmission. Cold Spring Harb. Perspect. Med. **2**(11) [Online] doi: 10.1101/cshperspect.a006965

26. **Coffin J, Swanstrom R**. 2013. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. Cold Spring Harb. Perspect. Med. **3**(1) [Online] doi: 10.1101/cshperspect.a012526

27. **Yao J.** 2009. Laboratory diagnosis of HIV infection: HIV markers during early infection. MAYO Clinic. [Online] URL:http://www.mayomedicallaboratories.com/articles/hottopics/transcripts/2009/2009-10a-hiv/2009-10a-hiv.html

28. **WHO**. 2007. WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children. WHO Press

29. **International Aids Society Scientific Working Group on HIV Cure, Deeks SG, Autran B, Berkhout B, Benkirane M, Cairns S, Chomont N, Chun TW, Churchill M, Di Mascio M, Katlama C, Lafeuillade A, Landay A, Lederman M, Lewin SR, Maldarelli F, Margolis D, Markowitz M, Martinez-Picado J, Mullins JI, Mellors J, Moreno S, O'Doherty U, Palmer S, Penicaud MC, Peterlin M, Poli G, Routy JP, Rouzioux C, Silvestri G, Stevenson M, Telenti A, Van Lint C, Verdin E, Woolfrey A, Zaia J, Barre-Sinoussi F.** 2012. Towards an HIV cure: a global scientific strategy. Nat. Rev. Immunol. **12**(8):607-614

30. **Sharp PM, Hahn BH**. 2011. Origins of HIV and the AIDS pandemic. Cold Spring Harb. Perspect. Med. [Online] doi: 10.1101/cshperspect.a006841

31. **Vallari A, Holzmayer V, Harris B, Yamaguchi J, Ngansop C, Makamche F, Mbanya D, Kaptue L, Ndembi N, Gurtler L, Devare S, Brennan CA.** 2011. Confirmation of putative HIV-1 group P in Cameroon. J. Virol. **85**(3):1403-1407

32. **Reeves JD, Doms RW.** 2002. Human immunodeficiency virus type 2. J. Gen. Virol. **83**(6):1253-1265

33. **Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, Muyembe JJ, Kabongo JM, Kalengayi RM, Van Marck E, Gilbert MT, Wolinsky SM.** 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature **455**(7213):661-664

34. **Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme AM.** 2003. Tracing the origin and history of the HIV-2 epidemic. Proc. Natl. Acad. Sci. USA **100**(11):6588-6592

35. **Los Alamos National Laboratory.** 2014. Landmarks of the HIV genome. Los Alamos National Laboratory. [Online] URL:http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html

36. **Karn J, Stoltzfus CM**. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb. Perspect. Med. **2**(2) [Online] doi: 10.1101/cshperspect.a006916

37. **Ivan Konstantinov YS, Alexander Kovalevsky, Yegor Voronin.** 2011. Human Immunodeficiency Virus Model. Visual Sciences.

38. **Wilen CB, Tilton JC, Doms RW.** 2012. Molecular mechanisms of HIV entry. Adv. Exp. Med. Biol. **726**:223-242

39. **Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI.** 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. Proc. Natl. Acad. Sci. USA **93**(7):3099-3104

40. **Hu WS, Hughes SH**. 2012. HIV-1 reverse transcription. Cold Spring Harb. Perspect. Med. **2**(10) [Online] doi: 10.1101/cshperspect.a006882

41. **Wilen CB, Tilton JC, Doms RW**. 2012. HIV: cell binding and entry. Cold Spring Harb. Perspect. Med. [Online] doi: 10.1101/cshperspect.a006866

42. **Craigie R, Bushman FD**. 2012. HIV DNA integration. Cold Spring Harb. Perspect. Med. **2**(7) [Online] doi: 10.1101/cshperspect.a006890

43. **Sundquist WI, Krausslich H-G**. 2012. HIV-1 assembly, budding and maturation. Cold Spring Harb. Perspect. Med. [Online] doi: 10.1101/cshperspect.a006924

44. **Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC.** 1997. Atomic structure of the ectodomain from HIV-1 gp41. Nature **387**(6631):426-430

45. **Shattock RJ, Moore JP.** 2003. Inhibiting sexual transmission of HIV-1 infection. Nat. Rev. Microbiol. **1**(1):25-34

46. **Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, Axel R.** 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune-system and the brain. Cell **47**(3):333-348

47. **McDougal JS, Kennedy MS, Sligh JM, Cort SP, Mawle A, Nicholson JKA.** 1986. Binding of HTLV-III/LAV to T4+ T-Cells by a complex of the 110k viral protein and the T4 molecule. Science **231**(4736):382-385

48.   **Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA.** 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature **393**(6686):648-659

49.   **Feng Y, Broder CC, Kennedy PE, Berger EA.** 1996. HIV-1 entry cofactor: Functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science **272**(5263):872-877

50.   **Dragic T, Litwin V, Allaway GP, Martin SR, Huang YX, Nagashima KA, Cayanan C, Maddon PJ, Koup RA, Moore JP, Paxton WA.** 1996. HIV-1 entry into CD4(+) cells is mediated by the chemokine receptor CC-CKR-5. Nature **381**(6584):667-673

51.   **Doranz BJ, Rucker J, Yi YJ, Smyth RJ, Samson M, Peiper SC, Parmentier M, Collman RG, Doms RW.** 1996. A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. Cell **85**(7):1149-1158

52.   **Deng HK, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, DiMarzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR, Landau NR.** 1996. Identification of a major co-receptor for primary isolates of HIV-1. Nature **381**(6584):661-666

53.   **Choe H, Farzan M, Sun Y, Sullivan N, Rollins B, Ponath PD, Wu LJ, Mackay CR, LaRosa G, Newman W, Gerard N, Gerard C, Sodroski J.** 1996. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. Cell **85**(7):1135-1148

54.   **Alkhatib G, Combadiere C, Broder CC, Feng Y, Kennedy PE, Murphy PM, Berger EA.** 1996. CC CKR5: a RANTES, MIP-1 alpha, MIP-1 beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science **272**(5270):1955-1958

55.   **Chan DC, Fass D, Berger JM, Kim PS.** 1997. Core structure of gp41 from the HIV envelope glycoprotein. Cell **89**(2):263-273

56.   **Melikyan GB**. 2008. Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm. Retrovirology **5**(111) [Online] doi: 10.1186/1742-4690-5-111

57.   **Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping L-H, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM.** 2008. Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. USA **105**(21):7552-7557

58.   **Regoes RR, Bonhoeffer S.** 2005. The HIV coreceptor switch: a population dynamical perspective. Trends Microbiol. **13**(6):269-277

59.   **Miyauchi K, Kim Y, Latinovic O, Morozov V, Melikyan GB.** 2009. HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. Cell **137**(3):433-444

60.   **Jolly C.** 2010. T cell polarization at the virological synapse. Viruses **2**(6):1261-1278

61. **Sattentau QJ.** 2010. Cell-to-cell spread of retroviruses. Viruses **2**(6):1306-1321

62. **Wild C, Greenwell T, Matthews T.** 1993. A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virus-mediated cell-cell fusion. AIDS Res. Hum. Retroviruses **9**(11):1051-1053

63. **Dorr P, Westby M, Dobbs S, Griffin P, Irvine B, Macartney M, Mori J, Rickett G, Smith-Burchnell C, Napier C, Webster R, Armour D, Price D, Stammen B, Wood A, Perros M.** 2005. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. Antimicrob. Agents Chemother. **49**(11):4721-4732

64. **Westby M, Lewis M, Whitcomb J, Youle M, Pozniak AL, James IT, Jenkins TM, Perros M, van der Ryst E.** 2006. Emergence of CXCR4-using human immunodeficiency virus type 1 (HIV-1) variants in a minority of HIV-1-infected patients following treatment with the CCR5 antagonist maraviroc is from a pretreatment CXCR4-using virus reservoir. J. Virol. **80**(10):4909-4920

65. **McDonald D, Vodicka MA, Lucero G, Svitkina TM, Borisy GG, Emerman M, Hope TJ.** 2002. Visualization of the intracellular behavior of HIV in living cells. J. Cell Biol. **159**(3):441-452

66. **Arhel N, Genovesio A, Kim KA, Miko S, Perret E, Olivo-Marin JC, Shorte S, Charneau P.** 2006. Quantitative four-dimensional tracking of cytoplasmic and nuclear HIV-1 complexes. Nat. Methods **3**(10):817-824

67. **Gaudin R, de Alencar BC, Arhel N, Benaroch P.** 2013. HIV trafficking in host cells: motors wanted! Trends Cell Biol. **23**(12):652-662

68. **Sluis-Cremer N, Arion D, Abram ME, Parniak MA.** 2004. Proteolytic processing of an HIV-1 pol polyprotein precursor: insights into the mechanism of reverse transcriptase p66/p51 heterodimer formation. Int. J. Biochem. Cell Biol. **36**(9):1836-1847

69. **Li S, Hill CP, Sundquist WI, Finch JT.** 2000. Image reconstructions of helical assemblies of the HIV-1CA protein. Nature **407**(6802):409-413

70. **Briggs JAG, Simon MN, Gross I, Krausslich H-G, Fuller SD, Vogt VM, Johnson MC.** 2004. The stoichiometry of Gag protein in HIV-1. Proc. Natl. Acad. Sci. USA **11**(7):672-675

71. **Arhel N.** 2010. Revisiting HIV-1 uncoating. Retrovirology **7**(96) [Online] doi: 10.1186/1742-4690-7-96

72. **Kleiman L, Jones CP, Musier-Forsyth K.** 2010. Formation of the tRNALys packaging complex in HIV-1. FEBS Lett. **584**(2):359-365

73. **Sarafianos SG, Das K, Tantillo C, Clark AD, Jr., Ding J, Whitcomb JM, Boyer PL, Hughes SH, Arnold E.** 2001. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. EMBO J. **20**(6):1449-1461

74. **Thomas JA, Gorelick RJ.** 2008. Nucleocapsid protein function in early infection processes. Virus Res. **134**(1-2):39-63

75. **Wu X, Liu H, Xiao H, Conway JA, Hehl E, Kalpana GV, Prasad V, Kappes JC.** 1999. Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions

with the nucleoprotein reverse transcription complex. J. Virol. **73**(3):2126-2135

76. **Engelman A, Cherepanov P.** 2012. The structural biology of HIV-1: mechanistic and therapeutic insights. Nat. Rev. Microbiol. **10**(4):279-290

77. **Arts EJ, Hazuda DJ**. 2012. HIV-1 antiretroviral drug therapy. Cold Spring Harb. Perspect. Med. **2**(4) [Online] doi: 10.1101/cshperspect.a007161

78. **Pornillos O, Ganser-Pornillos BK, Yeager M.** 2011. Atomic-level modelling of the HIV capsid. Nature **469**(7330):424-427

79. **Schaller T, Ocwieja KE, Rasaiyaah J, Price AJ, Brady TL, Roth SL, Hue S, Fletcher AJ, Lee K, KewalRamani VN, Noursadeghi M, Jenner RG, James LC, Bushman FD, Towers GJ.** 2011. HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. PLoS Path. **7**(12):e1002439

80. **Grewe C, Beck A, Gelderblom HR.** 1990. HIV: early virus-cell interactions. J. Acquir. Immune Defic. Syndr. **3**(10):965-974

81. **Fassati A, Goff SP.** 2001. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. J. Virol. **75**(8):3626-3635

82. **Miller MD, Farnet CM, Bushman FD.** 1997. Human immunodeficiency virus type 1 preintegration complexes: Studies of organization and composition. J. Virol. **71**(7):5382-5390

83. **Matreyek KA, Yucel SS, Li X, Engelman A.** 2013. Nucleoporin Nup153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. PLoS Path. **9**(10):e1003693

84. **Di Nunzio F, Fricke T, Miccio A, Valle-Casuso JC, Perez P, Souque P, Rizzi E, Severgnini M, Mavilio F, Charneau P, Diaz-Griffero F.** 2013. Nup153 and Nup98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. Virology **440**:8-18

85. **Aiken C.** 2009. Cell-free assays for HIV-1 uncoating, p. 41-53, Methods Mol. Biol., vol. 485.

86. **Forshey BM, von Schwedler U, Sundquist WI, Aiken C.** 2002. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. J. Virol. **76**(11):5667-5677

87. **Shah VB, Shi J, Hout DR, Oztop I, Krishnan L, Ahn J, Shotwell MS, Engelman A, Aiken C.** 2013. The host proteins transportin SR2/TNPO3 and cyclophilin A exert opposing effects on HIV-1 uncoating. J. Virol. **87**(1):422-432

88. **Bichel K, Price AJ, Schaller T, Towers GJ, Freund SM, James LC**. 2013. HIV-1 capsid undergoes coupled binding and isomerization by the nuclear pore protein NUP358. Retrovirology **10**(81) [Online] doi: 10.1186/1742-4690-10-81

89. **Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, Sundquist WI, Hill CP.** 1996. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. Cell **87**(7):1285-1294

90. **Fricke T, Brandariz-Nunez A, Wang X, Smith AB, 3rd, Diaz-Griffero F.** 2013. Human cytosolic extracts stabilize the HIV-1 core. J. Virol. **87**(19):10587-10597

91. **Xu H, Franks T, Gibson G, Huber K, Rahm N, De Castillia CS, Luban J, Aiken C, Watkins S, Sluis-Cremer N, Ambrose Z**. 2013. Evidence for biphasic uncoating during HIV-1 infection from a novel imaging assay. Retrovirology **10**:70 [Online] doi: 10.1186/1742-4690-10-70

92. **Rasaiyaah J, Tan CP, fletcher AJ, Price AJ, Blondeau C, Hilditch L, Jacques DA, Selwood DL, James LC, Noursadeghi M, Towers GJ.** 2013. HIV-1 evades innate immune recognition through specific cofactor recruitment. Nature **503**(7476):402-405

93. **Lahaye X, Satoh T, Gentili M, Cerboni S, Conrad C, Hurbain I, El Marjou A, Lacabaratz C, Lelievre JD, Manel N.** 2013. The capsids of HIV-1 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in dendritic cells. Immunity **39**(6):1132-1142

94. **Sodeik B, Ebersold MW, Helenius A.** 1997. Microtubule-mediated transport of incoming herpes simplex virus 1 capsids to the nucleus. J. Cell Biol. **136**(5):1007-1021

95. **Ojala PM, Sodeik B, Ebersold MW, Kutay U, Helenius A.** 2000. Herpes simplex virus type 1 entry into host cells: reconstitution of capsid binding and uncoating at the nuclear pore complex in vitro. Mol. Cell. Biol. **20**(13):4922-4931

96. **Batterson W, Furlong D, Roizman B.** 1983. Molecular genetics of herpes simplex virus: further characterization of a temperature-sensitive mutant defective in release of viral DNA and in other stages of the viral reproductive cycle. J. Virol. **45**(1):397-407

97. **Morgan C, Rosenkranz HS, Mednis B.** 1969. Structure and development of viruses as observed in the electron microscope. V. Entry and uncoating of adenovirus. J. Virol. **4**(5):777-796

98. **Greber UF, Suomalainen M, Stidwill RP, Boucke K, Ebersold MW, Helenius A.** 1997. The role of the nuclear pore complex in adenovirus DNA entry. EMBO J. **16**(19):5998-6007

99. **Summers MD.** 1971. Electron microscopic observations on granulosis virus entry, uncoating and replication processes during infection of the midgut cells of Trichoplusia ni. J. Ultrastruct. Res. **35**(5):606-625

100. **Arhel NJ, Souquere-Besse S, Munier S, Souque P, Guadagnini S, Rutherford S, Prévost M-C, Allen TD, Charneau P.** 2007. HIV-1 DNA flap formation promotes uncoating of the pre-integration complex at the nuclear pore. EMBO J. **26**(12):3025-3037

101. **Lelek M, Di Nunzio F, Henriques R, Charneau P, Arhel N, Zimmer C.** 2012. Superresolution imaging of HIV in infected cells with FlAsH-PALM. Proc. Natl. Acad. Sci. USA **109**(22):8564-8569

102. **Thomas JA, Ott DE, Gorelick RJ.** 2007. Efficiency of human immunodeficiency virus type 1 postentry infection processes: evidence against disproportionate numbers of defective virions. J. Virol. **81**(8):4367-4370

103. **Klarmann GJ, Schauber CA, Preston BD.** 1993. Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during

polymerization of HIV-1 sequences in vitro. J. Biol. Chem. **268**:9793-9802

104. **Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J.** 2004. The cytoplasmic body component TRIM5[alpha] restricts HIV-1 infection in Old World monkeys. Nature **427**(6977):848-853

105. **Yang Y, Fricke T, Diaz-Griffero F.** 2013. Inhibition of reverse transcriptase activity increases stability of the HIV-1 core. J. Virol. **87**(1):683-687

106. **Hulme AE, Hope TJ.** 2014. The cyclosporin A washout assay to detect HIV-1 uncoating in infected cells. Methods Mol. Biol. **1087**:37-46

107. **Hulme AE, Perez O, Hope TJ.** 2011. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. Proc. Natl. Acad. Sci. USA **108**(24):9975-9980

108. **Yang R, Shi J, Byeon I-JL, Ahn J, Sheehan JH, Meiler J, Gronenborn AM, Aiken C**. 2012. Second-site suppressors of HIV-1 capsid mutations: restoration of intracellular activities without correction of intrinsic capsid stability defects. Retrovirology **9**:30 [Online] doi: 10.1186/1742-4690-9-30

109. **Lukic Z, Dharan A, Fricke T, Diaz-Griffero F, Campbell EM.** 2014. HIV-1 uncoating is facilitated by dynein and kinesin 1. J. Virol. **88**(23):13613-13625

110. **Braaten D, Franke EK, Luban J.** 1996. Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. J. Virol. **70**(6):3551-3560

111. **Blair WS, Pickford C, Irving SL, Brown DG, Anderson M, Bazin R, Cao J, Ciaramella G, Isaacson J, Jackson L, Hunt R, Kjerrstrom A, Nieman JA, Patick AK, Perros M, Scott AD, Whitby K, Wu H, Butler SL.** 2010. HIV capsid is a tractable target for small molecule therapeutic intervention. PLoS Path. **6**(12):e1001220

112. **De Iaco A, Luban J**. 2014. Cyclophilin A promotes HIV-1 reverse transcription but its effect on transduction correlates best with its effect on nuclear entry of viral cDNA. Retrovirology **11**(11) [Online] doi: 10.1186/1742-4690-11-11

113. **Shi J, Zhou J, Shah VB, Aiken C, Whitby K.** 2011. Small-Molecule Inhibition of Human Immunodeficiency Virus Type 1 Infection by Virus Capsid Destabilization. J. Virol. **85**(1):542-549

114. **Stremlau M, Perron MJ, Lee M, Song B, Javanbakht H, Diaz-Griffero F, Anderson DJ, Sundquist WI, Sodroski J.** 2006. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5 restriction factor. Proc. Natl. Acad. Sci. USA **103**(14):5514-5519

115. **Black LR, Aiken C.** 2010. TRIM5 alpha disrupts the structure of assembled HIV-1 capsid complexes in vitro. J. Virol. **84**(13):6564-6569

116. **Chen N, Zhou L, Gane P, Price A, Zufferey M, Luban J, Fassati A**. 2013. HIV-1 uncoating in human CD4+ T cells: kinetic and functional analyses. Retrovirology **10**(13) [Online] doi: 10.1186/1742-4690-10-S1-P13

117. **Fricke T, White TE, Schulte B, de Souza Aranha Vieira DA, Dharan A, Campbell EM, Brandariz-Nunez A, Diaz-Griffero F**. 2014. MxB binds to

the HIV-1 core and prevents the uncoating process of HIV-1. Retrovirology **11**(68) [Online] doi: 10.1186/s12977-014-0068-x

118. **Hulme AE, Kelley Z, Okocha Z, Hope TJ.** 2014. Identification of capsid mutations that alter the rate of HIV-1 uncoating in infected cells. J. Virol. **89**(1):643-651

119. **Hoelz A, Debler EW, Blobel G.** 2011. The structure of the nuclear pore complex. Annu. Rev. Biochem. **80**:613-643

120. **Strambio-De-Castillia C, Niepel M, Rout MP.** 2010. The nuclear pore complex: bridging nuclear transport and gene regulation. Nat. Rev. Mol. Cell Biol. **11**(7):490-501

121. **Frey S, Richter RP, Gorlich D.** 2006. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. Science **314**(5800):815-817

122. **Fassati A**. 2006. HIV infection of non-dividing cells: a divisive problem. Retrovirology **3**(74) [Online] doi: 10.1186/1742-4690-3-74

123. **Yamashita M, Emerman M.** 2005. The cell cycle independence of HIV infections is not determined by known karyophilic viral elements. PLoS Path. **1**(3):e18

124. **Yamashita M, Emerman M.** 2004. Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. J. Virol. **78**(11):5670-5678

125. **Yamashita M, Perez O, Hope TJ, Emerman M.** 2007. Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. PLoS Path. **3**(10):1502-1510

126. **Farnet CM, Haseltine WA.** 1990. Integration of Human-Immunodeficiency-Virus Type-1 DNA Invitro. Proc. Natl. Acad. Sci. USA **87**(11):4164-4168

127. **Bushman FD, Craigie R.** 1991. Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. Proc. Natl. Acad. Sci. USA **88**(4):1339-1343

128. **Craigie R, Fujiwara T, Bushman F.** 1990. The IN protein of Moloney Murine Leukemia-Virus processes the viral-DNA ends and accomplishes their tntegration in vitro. Cell **62**(4):829-837

129. **Bushman FD, Fujiwara T, Craigie R.** 1990. Retroviral DNA Integration Directed by Hiv Integration Protein Invitro. Science **249**(4976):1555-1558

130. **Craigie R.** 2012. The molecular biology of HIV integrase. Mol. Immunol. **7**(7):679-686

131. **Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P.** 2010. Retroviral intasome assembly and inhibition of DNA strand transfer. Nature **464**(7286):232-108

132. **Maertens GN, Hare S, Cherepanov P.** 2010. The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature **468**(7321):326-U217

133. **Summa V, Petrocchi A, Bonelli F, Crescenzi B, Donghi M, Ferrara M, Fiore F, Gardelli C, Gonzalez Paz O, Hazuda DJ, Jones P, Kinzel O, Laufer R, Monteagudo E, Muraglia E, Nizi E, Orvieto F, Pace P, Pescatore G, Scarpelli R, Stillmock K, Witmer MV, Rowley M.** 2008. Discovery of raltegravir, a potent, selective orally bioavailable HIV-

integrase inhibitor for the treatment of HIV-AIDS infection. J. Med. Chem. **51**(18):5843-5855

134. **Meyerhans A, Breinig T, Vartanian JP, Wain-Hobson S.** 2003. Forms and function of intracellular HIV DNA, p. 14-21. *In* Leitner T, Foley B, Hahn B, Marx P, McCutchan F, Mellors J, Wolinsky S, Korber B (ed.), HIV Sequence Compendium 2003. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM.

135. **Jeanson L, Subra F, Vaganay S, Hervy M, Marangoni E, Bourhis J, Mouscadet JF.** 2002. Effect of Ku80 depletion on the preintegrative steps of HIV-1 replication in human cells. Virology **300**(1):100-108

136. **Sloan RD, Wainberg MA**. 2011. The role of unintegrated DNA in HIV infection. Retrovirology **8**(52) [Online] doi: 10.1186/1742-4690-8-52

137. **Holman AG, Coffin JM.** 2005. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. Proc. Natl. Acad. Sci. USA **102**(17):6103-6107

138. **Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD.** 2007. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res. **17**(8):1186-1194

139. **Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XQH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang JH, Miklos GLG, Nelson C, Broder S, Clark AG, Nadeau C, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng ZM, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge WM, Gong FC, Gu ZP, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke ZX, Ketchum KA, Lai ZW, Lei YD, Li ZY, Li JY, Liang Y, Lin XY, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue BX, Sun JT, Wang ZY, Wang AH, Wang X, Wang J, Wei MH, Wides R, Xiao CL, Yan CH, Yao A, Ye J, Zhan M, Zhang WQ, Zhang HY, Zhao Q, Zheng LS, Zhong F, Zhong WY, Zhu SPC, Zhao SY, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An HJ, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Glueksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R,**

Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi HY, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays AD, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu XJ, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen MY, Wu D, Wu M, Xia A, Zandieh A, Zhu XH.** 2001. The sequence of the human genome. Science **291**(5507):1304

140. **Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F.** 2002. HIV-1 integration in the human genome favors active genes and local hotspots. Cell **110**(4):521-529

141. **Han Y, Lassen K, Monie D, Sedaghat AR, Shimoji S, Liu X, Pierson TC, Margolick JB, Siliciano RF, Siliciano JD.** 2004. Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. J. Virol. **78**(12):6122-6133

142. **Barr SD, Ciuffi A, Leipzig J, Shinn P, Ecker JR, Bushman FD.** 2006. HIV integration site selection: targeting in macrophages and the effects of different routes of viral entry. Mol. Ther. **14**(2):218-225

143. **Stevens SW, Griffith JD.** 1996. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. J. Virol. **70**(9):6459-6462

144. **Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA.** 2012. The accessible chromatin landscape of the human genome. Nature **489**(7414):75-82

145. **Deaton AM, Bird A.** 2011. CpG islands and the regulation of transcription. Genes Dev. **25**(10):1010-1022

146. **Deininger P.** 2011. Alu elements: know the SINEs. Genome biology **12**(12):236

147. **Versteeg R, van Schaik BD, van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ, van Kampen AH.** 2003. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. Genome Res. **13**(9):1998-2004

148. **Ge H, Si YZ, Roeder RG.** 1998. Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. EMBO J. **17**(22):6723-6729

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

150. **Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, Tempe D, Vandekerckhove L, Moisant F, Ben-Slama L, Witvrouw M, Christ F, Rain JC, Dargemont C, Debyser Z, Benarous R.** 2005. Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. J. Biol. Chem. **280**(27):25517-25523

151. **Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, Debyser Z, Engelborghs Y.** 2003. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J. Biol. Chem. **278**(35):33528-33539

152. **Cherepanov P, Ambrosio ALB, Rahman S, Ellenberger T, Engelman A.** 2005. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc. Natl. Acad. Sci. USA **102**(48):17308-17313

153. **Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, Ecker JR, Bushman F.** 2005. A role for LEDGF/p75 in targeting HIV DNA integration. Nat. Med. **11**(12):1287-1289

154. **Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH, Teo W, Poeschla EM.** 2006. An essential role for LEDGF/p75 in HIV integration. Science **314**(5798):461-464

155. **Shun M-C, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, Kellam P, Cherepanov P, Engelman A.** 2007. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev. **21**(14):1767-1778

156. **Christ F, Voet A, Marchand A, Nicolet S, Desimmie BA, Marchand D, Bardiot D, Van der Veken NJ, Van Remoortel B, Strelkov SV, De Maeyer M, Chaltin P, Debyser Z.** 2010. Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nat. Chem. Biol. **6**(6):442-448

157. **Ciuffi A, Diamond TL, Hwang Y, Marshall HM, Bushman FD.** 2006. Modulating target site selection during human immunodeficiency virus DNA integration in vitro with an engineered tethering factor. Hum. Gene Ther. **17**(9):960-967

158. **Gijsbers R, Ronen K, Vets S, Malani N, De Rijck J, McNeely M, Bushman FD, Debyser Z.** 2010. LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. Mol. Ther. **18**(3):552-560

159. **Ocwieja KE, Brady TL, Ronen K, Huegel A, Roth SL, Schaller T, James LC, Towers GJ, Young JAT, Chanda SK, König R, Malani N, Berry CC, Bushman FD.** 2011. HIV integration targeting: a pathway involving transportin-3 and the nuclear pore protein RanBP2. PLoS Path. **7**(3):e1001313

160. **Carteau S, Hoffmann C, Bushman F.** 1998. Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. J. Virol. **72**(5):4005-4014

161. **Pruss D, Bushman FD, Wolffe AP.** 1994. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc. Natl. Acad. Sci. USA **91**(13):5913-5917

162. **Pruss D, Reeves R, Bushman FD, Wolffe AP.** 1994. The influence of DNA and nucleosome structure on integration events directed by HIV integrase. J. Biol. Chem. **269**(40):25031-25041

163. **Berry C, Hannenhalli S, Leipzig J, Bushman FD.** 2006. Selection of target sites for mobile DNA integration in the human genome. PLoS Comput. Biol. **2**(11):e157

164. **Lewinski MK, Bisgrove D, Shinn P, Chen H, Hoffmann C, Hannenhalli S, Verdin E, Berry CC, Ecker JR, Bushman FD.** 2005. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. J. Virol. **79**(11):6610-6619

165. **Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR, Bushman FD.** 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. **2**(8):e234

166. **Jordan A.** 2001. The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. EMBO J. **20**(7):1726-1738

167. **Ruelas DS, Greene WC.** 2013. An integrated overview of HIV-1 latency. Cell **155**(3):519-529

168. **Han Y, Lin YB, An W, Xu J, Yang HC, O'Connell K, Dordai D, Boeke JD, Siliciano JD, Siliciano RF.** 2008. Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. Cell Host Microbe **4**(2):134-146

169. **Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, Hannenhalli S, Hoffmann C.** 2005. Genome-wide analysis of retroviral DNA integration. Nat. Rev. Microbiol. **3**(11):848-858

170. **Sherrill-Mix S, Lewinski MK, Famiglietti M, Bosque A, Malani N, Ocwieja KE, Berry CC, Looney D, Shan L, Agosto LM, Pace MJ, Siliciano RF, O'Doherty U, Guatelli J, Planelles V, Bushman FD**. 2013. HIV latency and integration site placement in five cell-based models. Retrovirology **10**(90) [Online] doi: 10.1186/1742-4690-10-90

171. **Stoltzfus CM, Madsen JM.** 2006. Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. Curr. HIV Res. **4**(1):43-55

172. **Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH.** 2010. Crystal structure of HIV-1 Tat complexed with human P-TEFb. Nature **465**(7299):747-751

173. **Daugherty MD, Liu B, Frankel AD.** 2010. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. Nat. Struct. Mol. Biol. **17**(11):1337-1342

174. **Kutluay SB, Bieniasz PD.** 2010. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. PLoS Path. **6**(11):e1001200

175. **Datta SA, Heinrich F, Raghunandan S, Krueger S, Curtis JE, Rein A, Nanda H.** 2011. HIV-1 Gag extension: conformational changes require

simultaneous interaction with membrane and nucleic acid. J. Mol. Biol. **406**(2):205-214

176. **Kutluay SB, Zang T, Blanco-Melo D, Powell C, Jannain D, Errando M, Bieniasz PD.** 2014. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. Cell **159**(5):1096-1109

177. **Zhou W, Resh MD.** 1996. Differential membrane binding of the human immunodeficiency virus type 1 matrix protein. J. Virol. **70**(12):8540-8548

178. **Saad JS, Miller J, Tai J, Kim A, Ghanam RH, Summers MF.** 2006. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. Proc. Natl. Acad. Sci. USA **103**(30):11364-11369

179. **Jouvenet N, Bieniasz PD, Simon SM.** 2008. Imaging the biogenesis of individual HIV-1 virions in live cells. Nature **454**(7201):236-240

180. **Tedbury PR, Freed EO.** 2014. The role of matrix in HIV-1 envelope glycoprotein incorporation. Trends Microbiol. **22**(7):372-378

181. **Briggs JA, Riches JD, Glass B, Bartonova V, Zanetti G, Krausslich HG.** 2009. Structure and assembly of immature HIV. Proc. Natl. Acad. Sci. USA **106**(27):11090-11095

182. **Kondo E, Mammano F, Cohen EA, Gottlinger HG.** 1995. The p6gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. J. Virol. **69**(5):2759-2764

183. **Carlton JG, Martin-Serrano J.** 2007. Parallels between cytokinesis and retroviral budding: a role for the ESCRT machinery. Science **316**(5833):1908-1912

184. **Lata S, Schoehn G, Jain A, Pires R, Piehler J, Gottlinger HG, Weissenhorn W.** 2008. Helical structures of ESCRT-III are disassembled by VPS4. Science **321**(5894):1354-1357

185. **Strack B, Calistri A, Craig S, Popova E, Gottlinger HG.** 2003. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. Cell **114**(6):689-699

186. **Gottlinger HG, Dorfman T, Sodroski JG, Haseltine WA.** 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. Proc. Natl. Acad. Sci. USA **88**(8):3195-3199

187. **Pornillos O, Alam SL, Rich RL, Myszka DG, Davis DR, Sundquist WI.** 2002. Structure and functional interactions of the Tsg101 UEV domain. EMBO J. **21**(10):2397-2406

188. **Votteler J, Sundquist WI.** 2013. Virus budding and the ESCRT pathway. Cell Host Microbe **14**(3):232-241

189. **McCullough J, Fisher RD, Whitby FG, Sundquist WI, Hill CP.** 2008. ALIX-CHMP4 interactions in the human ESCRT pathway. Proc. Natl. Acad. Sci. USA **105**(22):7687-7691

190. **Morita E, Sandrin V, McCullough J, Katsuyama A, Baci Hamilton I, Sundquist WI.** 2011. ESCRT-III protein requirements for HIV-1 budding. Cell Host Microbe **9**(3):235-242

191. **Van Engelenburg SB, Shtengel G, Sengupta P, Waki K, Jarnik M, Ablan SD, Freed EO, Hess HF, Lippincott-Schwartz J.** 2014.

Distribution of ESCRT machinery at HIV assembly sites reveals virus scaffolding of ESCRT subunits. Science **343**(6171):653-656

192. **Wyma DJ, Jiang J, Shi J, Zhou J, Lineberger JE, Miller MD, Aiken C.** 2004. Coupling of Human Immunodeficiency Virus Type 1 fusion to virion maturation: a novel role of the gp41 cytoplasmic tail. J. Virol. **78**(7):3429-3435

193. **Li F, Goila-Gaur R, Salzwedel K, Kilgore NR, Reddick M, Matallana C, Castillo A, Zoumplis D, Martin DE, Orenstein JM, Allaway GP, Freed EO, Wild CT.** 2003. PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. Proc. Natl. Acad. Sci. USA **100**(23):13555-13560

194. **Sawyer SL, Wu LI, Emerman M, Malik HS.** 2005. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proc. Natl. Acad. Sci. USA **102**(8):2832-2837

195. **Sayah DM, Sokolskaja E, Berthoux L, Luban J.** 2004. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature **430**(6999):569-573

196. **Wilson SJ, Webb BL, Ylinen LM, Verschoor E, Heeney JL, Towers GJ.** 2008. Independent evolution of an antiviral TRIMCyp in rhesus macaques. Proc. Natl. Acad. Sci. USA **105**(9):3557-3562

197. **Ganser-Pornillos BK, Chandrasekaran V, Pornillos O, Sodroski JG, Sundquist WI, Yeager M.** 2011. Hexagonal assembly of a restricting TRIM5alpha protein. Proc. Natl. Acad. Sci. USA **108**(2):534-539

198. **Wu X, Anderson JL, Campbell EM, Joseph AM, Hope TJ.** 2006. Proteasome inhibitors uncouple rhesus TRIM5α restriction of HIV-1 reverse transcription and infection. Proc. Natl. Acad. Sci. USA **103**(19):7465-7470

199. **Diaz-Griffero F, Kar A, Perron M, Xiang SH, Javanbakht H, Li X, Sodroski J.** 2007. Modulation of retroviral restriction and proteasome inhibitor-resistant turnover by changes in the TRIM5alpha B-box 2 domain. J. Virol. **81**(19):10362-10378

200. **Sastri J, Campbell EM.** 2011. Recent insights into the mechanism and consequences of TRIM5α retroviral restriction. AIDS Res. Hum. Retroviruses **27**:231-238

201. **Towers GJ.** 2007. The control of viral infection by tripartite motif proteins and cyclophilin A. Retrovirology **4**(40) [Online] doi: 10.1186/1742-4690-4-40

202. **Pertel T, Hausmann S, Morger D, Zuger S, Guerra J, Lascano J, Reinhard C, Santoni FA, Uchil PD, Chatel L, Bisiaux A, Albert ML, Strambio-De-Castillia C, Mothes W, Pizzato M, Grutter MG, Luban J.** 2011. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. Nature **472**(7343):361-365

203. **Grutter MG, Luban J.** 2012. TRIM5 structure, HIV-1 capsid recognition, and innate immune signaling. Curr. Opin. Virol. **2**(2):142-150

204. **Varthakavi V, Smith RM, Bour SP, Strebel K, Spearman P.** 2003. Viral protein U counteracts a human host cell restriction that inhibits HIV-1 particle production. Proc. Natl. Acad. Sci. USA **100**(25):15154-15159

205. **Neil SJ, Eastman SW, Jouvenet N, Bieniasz PD.** 2006. HIV-1 Vpu promotes release and prevents endocytosis of nascent retrovirus particles from the plasma membrane. PLoS Path. **2**(5):e39

206. **Neil SJ, Sandrin V, Sundquist WI, Bieniasz PD.** 2007. An interferon-alpha-induced tethering mechanism inhibits HIV-1 and Ebola virus particle release but is counteracted by the HIV-1 Vpu protein. Cell Host Microbe **2**(3):193-203

207. **Van Damme N, Goff D, Katsura C, Jorgenson RL, Mitchell R, Johnson MC, Stephens EB, Guatelli J.** 2008. The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. Cell Host Microbe **3**(4):245-252

208. **Neil SJ, Zang T, Bieniasz PD.** 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature **451**(7177):425-430

209. **Perez-Caballero D, Zang T, Ebrahimi A, McNatt MW, Gregory DA, Johnson MC, Bieniasz PD.** 2009. Tetherin inhibits HIV-1 release by directly tethering virions to cells. Cell **139**(3):499-511

210. **Yang H, Wang J, Jia X, McNatt MW, Zang T, Pan B, Meng W, Wang HW, Bieniasz PD, Xiong Y.** 2010. Structural insight into the mechanisms of enveloped virus tethering by tetherin. Proc. Natl. Acad. Sci. USA **107**(43):18428-18432

211. **Galao RP, Le Tortorec A, Pickering S, Kueck T, Neil SJ.** 2012. Innate sensing of HIV-1 assembly by Tetherin induces NFkappaB-dependent proinflammatory responses. Cell Host Microbe **12**(5):633-644

212. **Kluge SF, Mack K, Iyer SS, Pujol FM, Heigele A, Learn GH, Usmani SM, Sauter D, Joas S, Hotter D, Bibollet-Ruche F, Plenderleith LJ, Peeters M, Geyer M, Sharp PM, Fackler OT, Hahn BH, Kirchhoff F.** 2014. Nef proteins of epidemic HIV-1 group O strains antagonize human tetherin. Cell Host Microbe **16**(5):639-650

213. **Le Tortorec A, Neil SJ.** 2009. Antagonism to and intracellular sequestration of human tetherin by the human immunodeficiency virus type 2 envelope glycoprotein. J. Virol. **83**(22):11966-11978

214. **Sauter D, Schindler M, Specht A, Landford WN, Munch J, Kim KA, Votteler J, Schubert U, Bibollet-Ruche F, Keele BF, Takehisa J, Ogando Y, Ochsenbauer C, Kappes JC, Ayouba A, Peeters M, Learn GH, Shaw G, Sharp PM, Bieniasz P, Hahn BH, Hatziioannou T, Kirchhoff F.** 2009. Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. Cell Host Microbe **6**(5):409-421

215. **Sheehy AM, Gaddis NC, Choi JD, Malim MH.** 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature **418**(6898):646-650

216. **Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L.** 2003. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature **424**(6944):94-98

217. **Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D.** 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature **424**(6944):99-103

218. **Bishop KN, Verma M, Kim EY, Wolinsky SM, Malim MH.** 2008. APOBEC3G inhibits elongation of HIV-1 reverse transcripts. PLoS Path. **4**(12):e1000231

219. **Malim MH, Bieniasz PD**. 2012. HIV restriction factors and mechanisms of evasion. Cold Spring Harb. Perspect. Med. **2**(5) [Online] doi: 10.1101/cshperspect.a006940

220. **Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, Yu XF.** 2003. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science **302**(5647):1056-1060

221. **Kaushik R, Zhu X, Stranska R, Wu Y, Stevenson M.** 2009. A cellular restriction dictates the permissivity of nondividing monocytes/macrophages to lentivirus and gammaretrovirus infection. Cell Host Microbe **6**(1):68-80

222. **Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Segeral E, Yatim A, Emiliani S, Schwartz O, Benkirane M.** 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature **474**(7353):654-657

223. **Hrecka K, Hao C, Gierszewska M, Swanson SK, Kesik-Brodacka M, Srivastava S, Florens L, Washburn MP, Skowronski J.** 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. Nature **474**(7353):658-661

224. **Baldauf H-M, Pan X, Erikson E, Schmidt S, Daddacha W, Burggraf M, Schenkova K, Ambiel I, Wabnitz G, Gramberg T, Panitz S, Flory E, Landau NR, Sertel S, Rutsch F, Lasitschka F, Kim B, König R, Fackler OT, Keppler OT.** 2012. SAMHD1 restricts HIV-1 infection in resting CD4+ T cells. Nat. Med. **18**(11):1682-1687

225. **Goldstone DC, Ennis-Adeniran V, Hedden JJ, Groom HCT, Rice GI, Christodoulou E, Walker PA, Kelly G, Haire LF, Yap MW, de Carvalho LPS, Stoye JP, Crow YJ, Taylor IA, Webb M.** 2011. HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. Nature **480**(7377):379-382

226. **Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, Rice CM.** 2011. A diverse range of gene products are effectors of the type I interferon antiviral response. Nature **472**(7344):481-485

227. **Schwefel D, Groom HC, Boucherit VC, Christodoulou E, Walker PA, Stoye JP, Bishop KN, Taylor IA.** 2014. Structural basis of lentiviral subversion of a cellular protein degradation pathway. Nature **505**(7482):234-238

228. **Ayinde D, Casartelli N, Schwartz O.** 2012. Restricting HIV the SAMHD1 way: through nucleotide starvation. Nat. Rev. Microbiol. **10**(10):675-680

229. **Foster JL, Garcia JV**. 2008. HIV-1 Nef: at the crossroads. Retrovirology **5**(84) [Online] doi: 10.1186/1742-4690-5-84

230. **Gorry PR, McPhee DA, Verity E, Dyer WB, Wesselingh SL, Learmont J, Sullivan JS, Roche M, Zaunders JJ, Gabuzda D, Crowe SM, Mills J, Lewin SR, Brew BJ, Cunningham AL, Churchill MJ.** 2007. Pathogenicity and immunogenicity of attenuated, nef-deleted HIV-1 strains in vivo. Retrovirology **4**(66) [Online] doi: 10.1186/1742-4690-4-66

231. **Pizzato M, Helander A, Popova E, Calistri A, Zamborlini A, Palu G, Gottlinger HG.** 2007. Dynamin 2 is required for the enhancement of HIV-1 infectivity by Nef. Proc. Natl. Acad. Sci. USA **104**(16):6812-6817

232. **Chaudhuri R, Lindwasser OW, Smith WJ, Hurley JH, Bonifacino JS.** 2007. Downregulation of CD4 by human immunodeficiency virus type 1 Nef is dependent on clathrin and involves direct interaction of Nef with the AP2 clathrin adaptor. J. Virol. **81**(8):3877-3890

233. **Jia X, Singh R, Homann S, Yang H, Guatelli J, Xiong Y.** 2012. Structural basis of evasion of cellular adaptive immunity by HIV-1 Nef. Nat. Struct. Mol. Biol. **19**(7):701-706

234. **Magadan JG, Bonifacino JS.** 2012. Transmembrane domain determinants of CD4 Downregulation by HIV-1 Vpu. J. Virol. **86**(2):757-772

235. **Jowett JB, Planelles V, Poon B, Shah NP, Chen ML, Chen IS.** 1995. The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle. J. Virol. **69**(10):6304-6313

236. **Laguette N, Bregnard C, Hue P, Basbous J, Yatim A, Larroque M, Kirchhoff F, Constantinou A, Sobhian B, Benkirane M.** 2014. Premature activation of the SLX4 complex by Vpr promotes G2/M arrest and escape from innate immune sensing. Cell **156**(1):134-145

237. **de Noronha CM, Sherman MP, Lin HW, Cavrois MV, Moir RD, Goldman RD, Greene WC.** 2001. Dynamic disruptions in nuclear envelope architecture and integrity induced by HIV-1 Vpr. Science **294**(5544):1105-1108

238. **Stewart SA, Poon B, Jowett JB, Xie Y, Chen IS.** 1999. Lentiviral delivery of HIV-1 Vpr protein induces apoptosis in transformed cells. Proc. Natl. Acad. Sci. USA **96**(21):12039-12043

239. **Felzien LK, Woffendin C, Hottiger MO, Subbramanian RA, Cohen EA, Nabel GJ.** 1998. HIV transcriptional activation by the accessory protein, VPR, is mediated by the p300 co-activator. Proc. Natl. Acad. Sci. USA **95**(9):5281-5286

240. **Popov S, Rexach M, Zybarth G, Reiling N, Lee MA, Ratner L, Lane CM, Moore MS, Blobel G, Bukrinsky M.** 1998. Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. EMBO J. **17**(4):909-917

241. **Lai MC, Lin RI, Tarn WY.** 2001. Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. Proc. Natl. Acad. Sci. USA **98**(18):10154-10159

242. **Christ F, Thys W, De Rijck J, Gijsbers R, Albanese A, Arosio D, Emiliani S, Rain J-C, Benarous R, Cereseto A, Debyser Z.** 2008. Transportin-SR2 imports HIV into the nucleus. Curr. Biol. **18**(16):1192-1202

243. **König R, Zhou Y, Elleder D, Diamond TL, Bonamy GMC, Irelan JT, Chiang C-y, Tu BP, De Jesus PD, Lilley CE, Seidel S, Opaluch AM, Caldwell JS, Weitzman MD, Kuhen KL, Bandyopadhyay S, Ideker T, Orth AP, Miraglia LJ, Bushman FD, Young JA, Chanda SK.** 2008. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell **135**(1):49-60

244. **Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ.** 2008. Identification of host proteins required for HIV infection through a functional genomic screen. Science **319**(5865):921-926

245. **Valle-Casuso JC, Di Nunzio F, Yang Y, Reszka N, Lienlaf M, Arhel N, Perez P, Brass AL, Diaz-Griffero F.** 2012. TNPO3 is required for HIV-1 replication after nuclear import but prior to integration and binds the HIV-1 core. J. Virol. **86**(10):5931-5936

246. **De Iaco A, Luban J**. 2011. Inhibition of HIV-1 infection by TNPO3 depletion is determined by capsid and detectable after viral cDNA enters the nucleus. Retrovirology **8**(98) [Online] doi: 10.1186/1742-4690-8-98

247. **De Iaco A, Santoni F, Vannier A, Guipponi M, Antonarakis S, Luban J**. 2013. TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm. Retrovirology **10**(20) [Online] doi: 10.1186/1742-4690-10-20

248. **Zhou L, Sokolskaja E, Jolly C, James W, Cowley SA, Fassati A.** 2011. Transportin 3 promotes a nuclear maturation step required for efficient HIV-1 integration. PLoS Path. **7**(8):e1002194

249. **Larue R, Gupta K, Wuensch C, Shkriabai N, Kessl JJ, Danhart E, Feng L, Taltynov O, Christ F, Van Duyne GD, Debyser Z, Foster MP, Kvaratskhelia M.** 2012. Interaction of the HIV-1 intasome with transportin 3 protein (TNPO3 or TRN-SR2). J. Biol. Chem. **287**(41):34044-34058

250. **Krishnan L, Matreyek KA, Oztop I, Lee K, Tipper CH, Li X, Dar MJ, KewalRamani VN, Engelman A.** 2010. The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to Human Immunodeficiency Virus Type 1 capsid and not integrase. J. Virol. **84**(1):397-406

251. **Thys W, De Houwer S, Demeulemeester J, Taltynov O, Vancraenenbroeck R, Gerard M, De Rijck J, Gijsbers R, Christ F, Debyser Z**. 2011. Interplay between HIV entry and transportin-SR2 dependency. Retrovirology **8**(7) [Online] doi: 10.1186/1742-4690-8-7

252. **Taltynov O, Demeulemeester J, Christ F, De Houwer S, Tsirkone VG, Gerard M, Weeks SD, Strelkov SV, Debyser Z.** 2013. Interaction of transportin-SR2 with Ras-related nuclear protein (Ran) GTPase. J. Biol. Chem. **288**(35):25603-25613

253. **De Houwer S, Demeulemeester J, Thys W, Taltynov O, Zmajkovicova K, Christ F, Debyser Z.** 2012. Identification of residues in the C-terminal domain of HIV-1 integrase that mediate binding to the transportin-SR2 protein. J. Biol. Chem. **287**(41):34059-34068

254. **Wu J, Matunis MJ, Kraemer D, Blobel G, Coutavas E.** 1995. Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. J. Biol. Chem. **270**(23):14209-14213

255. **Wälde S, Thakar K, Hutten S, Spillner C, Nath A, Rothbauer U, Wiemann S, Kehlenbach RH.** 2012. The nucleoporin Nup358/RanBP2 promotes nuclear import in a cargo‐ and transport receptor‐specific manner. Traffic **13**(2):218-233

256. **Hutten S, Walde S, Spillner C, Hauber J, Kehlenbach RH.** 2009. The nuclear pore component Nup358 promotes transportin-dependent nuclear import. Proc. Natl. Acad. Sci. USA **122**(8):1100-1110

257. **Zhang R, Mehla R, Chauhan A.** 2010. Perturbation of Host Nuclear Membrane Component RanBP2 Impairs the Nuclear Import of Human Immunodeficiency Virus -1 Preintegration Complex (DNA). PLoS One **5**(12):e15620

258. **Di Nunzio F, Danckaert A, Fricke T, Perez P, Fernandez J, Perret E, Roux P, Shorte S, Charneau P, Diaz-Griffero F, Arhel NJ.** 2012. Human nucleoporins promote HIV-1 docking at the nuclear pore, nuclear import and integration. PLoS One **7**(9):e46037

259. **Lee K, Ambrose Z, Martin TD, Oztop I, Mulky A, Julias JG, Vandegraaff N, Baumann JG, Wang R, Yuen W, Takemura T, Shelton K, Taniuchi I, Li Y, Sodroski J, Littman DR, Coffin JM, Hughes SH, Unutmaz D, Engelman A, KewalRamani VN.** 2010. Flexible use of nuclear import pathways by HIV-1. Cell Host Microbe **7**(3):221-233

260. **Meehan AM, Saenz DT, Guevera R, Morrison JH, Peretz M, Fadel HJ, Hamada M, van Deursen J, Poeschla EM.** 2014. A cyclophilin homology domain-independent role for Nup358 in HIV-1 infection. PLoS Path. **10**(2):e1003969

261. **Walther TC, Fornerod M, Pickersgill H, Goldberg M, Allen TD, Mattaj IW.** 2001. The nucleoporin Nup153 is required for nuclear pore basket formation, nuclear pore complex anchoring and import of a subset of nuclear proteins. EMBO J. **20**(20):5703-5714

262. **Vaquerizas JM, Suyama R, Kind J, Miura K, Luscbome NM, Akhtar A.** 2010. Nuclear pore proteins Nup153 and Megator define transcriptionally active regions in the *Drospophila* genome. PLoS Genet. **6**(2):e1000846

263. **Griffis ER.** 2004. Distinct functional domains within nucleoporins Nup153 and Nup98 mediate transcription-dependent mobility. Mol. Biol. Cell **15**(4):1991-2002

264. **Daigle N, Beaudouin J, Hartnell L, Imreh G, Hallberg E, Lippincott-Schwartz J, Ellenberg J.** 2001. Nuclear pore complexes form immobile networks and have a very low turnover in live mammalian cells. J. Cell Biol. **154**(1):71-84

265. **Ball JR, Ullman KS.** 2005. Versatility at the nuclear pore complex: lessons learned from the nucleoporin Nup153. Chromosoma **114**(5):319-330

266. **Matreyek KA, Engelman A.** 2011. The requirement for nucleoporin Nup153 during Human Immunodeficiency Virus type 1 Infection is determined by the viral capsid. J. Virol. **85**(15):7818-7827

267. **Woodward CL, Prakobwanakit S, Mosessian S, Chow SA.** 2009. Integrase interacts with nucleoporin Nup153 to mediate the nuclear import of Human Immunodeficiency Virus type 1. J. Virol. **83**(13):6522-6533

268. **Koh Y, Wu X, Ferris AL, Matreyek KA, Smith SJ, Lee K, KewalRamani VN, Hughes SH, Engelman A.** 2012. Differential effects of Human Immunodeficiency Virus type 1 capsid and cellular factors

Nucleoporin153 and LEDGF/p75 on the efficiency and specificity of viral DNA integration. J. Virol. **87**(1):648-658

269. **Ao Z, Danappa Jayappa K, Wang B, Zheng Y, Wang X, Peng J, Yao X.** 2012. Contribution of host nucleoporin 62 in HIV-1 integrase chromatin association and viral DNA integration. J. Biol. Chem. **287**(13):10544-10555

270. **Fassati A, Gorlich D, Harrison I, Zaytseva L, Mingot JM.** 2003. Nuclear import of HIV-1 intracellular reverse transcription complexes is mediated by importin 7. EMBO J. **22**(14):3675-3685

271. **Ao Z, Huang G, Yao H, Xu Z, Labine M, Cochrane AW, Yao X.** 2007. Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. J. Biol. Chem. **282**(18):13456-13467

272. **Zaitseva L, Cherepanov P, Leyens L, Wilson SJ, Rasaiyaah J, Fassati A**. 2009. HIV-1 exploits importin 7 to maximize nuclear import of its DNA genome. Retrovirology **6**(11) [Online] doi: 10.1186/1742-4690-6-11

273. **Wang P, Heitman J.** 2005. The cyclophilins. Genome Biol. **6**(226)

274. **Davis TL, Walker JR, Campagna-Slater V, Finerty PJ, Paramanathan R, Bernstein G, MacKenzie F, Tempel W, Ouyang H, Lee WH, Eisenmesser EZ, Dhe-Paganon S.** 2010. Structural and biochemical characterization of the human cyclophilin family of peptidyl-prolyl isomerases. PLoS Biol. **8**(7):e1000439

275. **Nigro P, Pompilio G, Capogrossi MC.** 2013. Cyclophilin A: a key player for human disease. Cell Death Dis. **4**:e888

276. **Handschumacher RE, Harding MW, Rice J, Drugge RJ, Speicher DW.** 1984. Cyclophilin: a specific cytosolic binding protein for cyclosporin A. Science **226**(4674):544-547

277. **Fischer G, Wittmann-Liebold B, Lang K, Kiefhaber T, Schmid FX.** 1989. Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins. Nature **337**(6206):476-478

278. **Takahashi N, Hayano T, Suzuki M.** 1989. Peptidyl-prolyl cis-trans isomerase is the cyclosporin A-binding protein cyclophilin. Nature **337**(6206):473-475

279. **Okamura H, Aramburu J, Garcia-Rodriguez C, Viola JP, Raghavan A, Tahiliani M, Zhang X, Qin J, Hogan PG, Rao A.** 2000. Concerted dephosphorylation of the transcription factor NFAT1 induces a conformational switch that regulates transcriptional activity. Mol. Cell **6**(3):539-550

280. **Jain J, McCaffrey PG, Miner Z, Kerppola TK, Lambert JN, Verdine GL, Curran T, Rao A.** 1993. The T-cell transcription factor NFATp is a substrate for calcineurin and interacts with Fos and Jun. Nature **365**(6444):352-355

281. **Jain J, McCaffrey PG, Valge-Archer VE, Rao A.** 1992. Nuclear factor of activated T cells contains Fos and Jun. Nature **356**(6372):801-804

282. **Liu J, Farmer JD, Jr., Lane WS, Friedman J, Weissman I, Schreiber SL.** 1991. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. Cell **66**(4):807-815

283. **Jin L, Harrison SC.** 2002. Crystal structure of human calcineurin complexed with cyclosporin A and human cyclophilin. Proc. Natl. Acad. Sci. USA **99**(21):13522-13526

284. **Huai Q, Kim HY, Liu Y, Zhao Y, Mondragon A, Liu JO, Ke H.** 2002. Crystal structure of calcineurin-cyclophilin-cyclosporin shows common but distinct recognition of immunophilin-drug complexes. Proc. Natl. Acad. Sci. USA **99**(19):12037-12042

285. **Hogan PG, Chen L, Nardone J, Rao A.** 2003. Transcriptional regulation by calcium, calcineurin, and NFAT. Genes Dev. **17**(18):2205-2232

286. **Beals CR, Clipstone NA, Ho SN, Crabtree GR.** 1997. Nuclear localization of NF-ATc by a calcineurin-dependent, cyclosporin-sensitive intramolecular interaction. Genes Dev. **11**(7):824-834

287. **Luban J, Bossolt KL, Franke EK, Kalpana GV, Goff SP.** 1993. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. Cell **73**(6):1067-1078

288. **Franke EK, Yuan HE, Luban J.** 1994. Specific incorporation of cyclophilin A into HIV-1 virions. Nature **372**(6504):359-362

289. **Thali M, Bukovsky A, Kondo E, Rosenwirth B, Walsh CT, Sodroski J, Gottlinger HG.** 1994. Functional association of cyclophilin A with HIV-1 virions. Nature **372**(6504):363-365

290. **Yoo S, Myszka DG, Yeh C, McMurray M, Hill CP, Sundquist WI.** 1997. Molecular recognition in the HIV-1 capsid/cyclophilin A complex. J. Mol. Biol. **269**(5):780-795

291. **Mikol V, Kallen J, Pflugl G, Walkinshaw MD.** 1993. X-ray structure of a monomeric cyclophilin A-cyclosporin A crystal complex at 2.1 A resolution. J. Mol. Biol. **234**(4):1119-1130

292. **Braaten D, Franke EK, Luban J.** 1996. Cyclophilin A is required for the replication of group M human immunodeficiency virus type 1 (HIV-1) and simian immunodeficiency virus SIV(CPZ)GAB but not group O HIV-1 or other primate immunodeficiency viruses. J. Virol. **70**(7):4220-4227

293. **Braaten D, Luban J.** 2001. Cyclophilin A regulates HIV-1 infectivity, as demonstrated by gene targeting in human T cells. EMBO J. **20**(6):1300-1309

294. **Towers GJ, Hatziioannou T, Cowan S, Goff SP, Luban J, Bieniasz PD.** 2003. Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. Nat. Med. **9**(9):1138-1143

295. **Hatziioannou T, Perez-Caballero D, Cowan S, Bieniasz PD.** 2005. Cyclophilin interactions with incoming Human Immunodeficiency Virus type 1 capsids with opposing effects on infectivity in human cells. J. Virol. **79**(1):176-183

296. **Yin L, Braaten D, Luban J.** 1998. Human immunodeficiency virus type 1 replication is modulated by host cyclophilin A expression levels. J. Virol. **72**(8):6430-6436

297. **Braaten D, Aberham C, Franke EK, Yin L, Phares W, Luban J.** 1996. Cyclosporine A-resistant human immunodeficiency virus type 1 mutants demonstrate that Gag encodes the functional target of cyclophilin A. J. Virol. **70**(8):5170

298. **Ylinen LMJ, Schaller T, Price A, Fletcher AJ, Noursadeghi M, James LC, Towers GJ.** 2009. Cyclophilin A levels dictate infection efficiency of

Human Immunodeficiency Virus type 1 capsid escape mutants A92E and G94D. J. Virol. **83**(4):2044-2047

299. **Song C, Aiken C.** 2007. Analysis of human cell heterokaryons demonstrates that target cell restriction of cyclosporine-resistant human immunodeficiency virus type 1 mutants is genetically dominant. J. Virol. **81**(21):11946-11956

300. **Shah VB, Aiken C.** 2014. Gene expression analysis of a panel of cell lines that differentially restrict HIV-1 CA mutants infection in a cyclophilin a-dependent manner. PLoS One **9**(3):e92724

301. **Matsuoka S, Dam E, Lecossier D, Clavel F, Hance AJ**. 2009. Modulation of HIV-1 infectivity and cyclophilin A-dependence by Gag sequence and target cell type. Retrovirology **6**(21) [Online] doi: 10.1186/1742-4690-6-21

302. **Berthoux L, Sebastian S, Sokolskaja E, Luban J.** 2005. Cyclophilin A is required for TRIM5{alpha}-mediated resistance to HIV-1 in Old World monkey cells. Proc. Natl. Acad. Sci. USA **102**(41):14849-14853

303. **Stremlau M, Song B, Javanbakht H, Perron M, Sodroski J.** 2006. Cyclophilin A: An auxiliary but not necessary cofactor for TRIM5α restriction of HIV-1. Virology **351**(1):112-120

304. **Sokolskaja E, Berthoux L, Luban J.** 2006. Cyclophilin A and TRIM5 independently regulate Human Immunodeficiency Virus type 1 infectivity in human cells. J. Virol. **80**(6):2855-2862

305. **Bosco DA, Eisenmesser EZ, Pochapsky S, Sundquist WI, Kern D.** 2002. Catalysis of cis/trans isomerization in native HIV-1 capsid by human cyclophilin A. Proc. Natl. Acad. Sci. USA **99**(8):5247-5252

306. **Lammers M, Neumann H, Chin JW, James LC.** 2010. Acetylation regulates cyclophilin A catalysis, immunosuppression and HIV isomerization. Nat. Chem. Biol. **6**(5):331-337

307. **Goujon C, Moncorge O, Bauby H, Doyle T, Ward CC, Schaller T, Hue S, Barclay WS, Schulz R, Malim MH.** 2013. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. Nature **502**(7472):559-562

308. **Kane M, Yadav SS, Bitzegeio J, Kutluay SB, Zang T, Wilson SJ, Schoggins JW, Rice CM, Yamashita M, Hatziioannou T, Bieniasz PD.** 2013. MX2 is an interferon-induced inhibitor of HIV-1 infection. Nature **502**(7472):563-566

309. **Liu Z, Pan Q, Ding S, Qian J, Xu F, Zhou J, Cen S, Guo F, Liang C.** 2013. The interferon-inducible MxB protein inhibits HIV-1 infection. Cell Host Microbe **14**(4):398-410

310. **Xiang K, Tong L, Manley JL.** 2014. Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. Mol. Cell. Biol. **34**(11):1894-1910

311. **Takagaki Y, Ryner LC, Manley JL.** 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. Genes Dev. **3**(11):1711-1724

312. **Ruegsegger U, Beyer K, Keller W.** 1996. Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. J. Biol. Chem. **271**(11):6107-6113

313. **Ruegsegger U, Blank D, Keller W.** 1998. Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. Mol. Cell **1**(2):243-253

314. **Kim S, Yamamoto J, Chen Y, Aida M, Wada T, Handa H, Yamaguchi Y.** 2010. Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. Genes Cells **15**(9):1003-1013

315. **Yang Q, Coseno M, Gilmartin GM, Doublié S.** 2011. Crystal structure of a human cleavage factor CFIm25/CFIm68/RNA complex provides an insight into poly(A) site recognition and RNA looping. Structure **19**(3):368-377

316. **Ruepp MD, Schumperli D, Barabino SM.** 2011. mRNA 3' end processing and more - multiple functions of mammalian cleavage factor I-68. WIRE RNA **2**(1):79-91

317. **Dettwiler S, Aringhieri C, Cardinale S, Keller W, Barabino SML.** 2004. Distinct sequence motifs within the 68-kDa subunit of cleavage factor I-m mediate RNA binding, protein-protein interactions, and subcellular localization. J. Biol. Chem. **279**(34):35788-35797

318. **Brown KM, Gilmartin GM.** 2003. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Mol. Cell **12**(6):1467-1476

319. **Awasthi S, Alwine JC.** 2003. Association of polyadenylation cleavage factor I with U1 snRNP. RNA **9**(11):1400-1409

320. **Ryan K.** 2007. Pre-mRNA 3' cleavage is reversibly inhibited in vitro by cleavage factor dephosphorylation. RNA Biol. **4**(1):26-33

321. **Cardinale S, Cisterna B, Bonetti P, Aringhieri C, Biggiogera M, Barabino SM.** 2007. Subnuclear localization and dynamics of the Pre-mRNA 3' end processing factor mammalian cleavage factor I 68-kDa subunit. Mol. Biol. Cell **18**(4):1282-1292

322. **Ruepp MD, Aringhieri C, Vivarelli S, Cardinale S, Paro S, Schumperli D, Barabino SM.** 2009. Mammalian pre-mRNA 3' end processing factor CF I m 68 functions in mRNA export. Mol. Biol. Cell **20**(24):5211-5223

323. **Guttinger S, Muhlhausser P, Koller-Eichhorn R, Brennecke J, Kutay U.** 2004. Transportin2 functions as importin and mediates nuclear import of HuR. Proc. Natl. Acad. Sci. USA **101**(9):2918-2923

324. **Lee BJ, Cansizoglu AE, Suel KE, Louis TH, Zhang Z, Chook YM.** 2006. Rules for nuclear localization sequence recognition by karyopherin beta 2. Cell **126**(3):543-558

325. **Brockman MA, Schneidewind A, Lahaie M, Schmidt A, Miura T, DeSouza I, Ryvkin F, Derdeyn CA, Allen S, Hunter E, Mulenga J, Goepfert PA, Walker BD, Allen TM.** 2007. Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on Human Immunodeficiency Virus type 1 Gag alter capsid interactions with cyclophilin A. J. Virol. **81**(22):12608-12618

326. **Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, Le Gall S, Rinaldo CR, Craggs SL, Allgaier RL, Power KA, Kuntzen T, Tung CS, LaBute MX, Mueller SM, Harrer T, McMichael AJ, Goulder PJR, Aiken C, Brander C, Kelleher AD, Allen TM.** 2007. Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is

associated with a dramatic reduction in Human Immunodeficiency Virus type 1 replication. J. Virol. **81**(22):12382-12393

327. **Lamorte L, Titolo S, Lemke CT, Goudreau N, Mercier JF, Wardrop E, Shah VB, von Schwedler UK, Langelier C, Banik SS, Aiken C, Sundquist WI, Mason SW.** 2013. Discovery of novel small-molecule HIV-1 replication inhibitors that stabilize capsid complexes. Antimicrob. Agents Chemother. **57**(10):4622-4631

328. **Yee JK, Miyanohara A, LaPorte P, Bouic K, Burns JC, Friedmann T.** 1994. A general method for the generation of high-titer, pantropic retroviral vectors: highly efficient infection of primary hepatocytes. Proc. Natl. Acad. Sci. USA **91**(20):9564-9568

329. **Ikeda Y, Ylinen LM, Kahar-Bador M, Towers GJ.** 2004. Influence of gag on human immunodeficiency virus type 1 species-specific tropism. J. Virol. **78**(21):11816-11822

330. **Bock M, Bishop KN, Towers G, Stoye JP.** 2000. Use of a transient assay for studying the genetic determinants of Fv1 restriction. J. Virol. **74**(16):7422-7430

331. **Towers G, Bock M, Martin S, Takeuchi Y, Stoye JP, Danos O.** 2000. A conserved mechanism of retrovirus restriction in mammals. Proc. Natl. Acad. Sci. USA **97**(22):12295-12299

332. **Demaison C, Parsley K, Brouns G, Scherr M, Battmer K, Kinnon C, Grez M, Thrasher AJ.** 2002. High-level transduction and gene expression in hematopoietic repopulating cells using a human immunodeficiency virus type 1-based lentiviral vector containing an internal spleen focus forming virus promoter. Hum. Gene Ther. **13**(7):803-813

333. **Neil S, Martin F, Ikeda Y, Collins M.** 2001. Postentry restriction to human immunodeficiency virus-based vector transduction in human monocytes. J. Virol. **75**(12):5448-5456

334. **Zhang F, Hatziioannou T, Perez-Caballero D, Derse D, Bieniasz PD.** 2006. Antiretroviral potential of human tripartite motif-5 and related proteins. Virology **353**(2):396-409

335. **Henning MS, Dubose BN, Burse MJ, Aiken C, Yamashita M.** 2014. In vivo functions of CPSF6 for HIV-1 as revealed by HIV-1 capsid evolution in HLA-B27-positive subjects. PLoS Path. **10**(1):e1003868

336. **Boyer HW, Roulland-Dussoix D.** 1969. A complementation analysis of the restriction and modification of DNA in Escherichia coli. J. Mol. Biol. **41**(3):459-472

337. **UnitProt.** 2014. Cleavage and polyadenylation specificity factor subunit 6. UniProt Consortium. [Online] URL:http://www.uniprot.org/uniprot/Q5ZL34

338. **Roe T, Reynolds TC, Yu G, Brown PO.** 1993. Integration of murine leukemia virus DNA depends on mitosis. EMBO J. **12**(5):2099-2108

339. **Shimodaira H.** 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. **51**(3):492-508

340. **Price AJ, Jacques DA, McEwan WA, Fletcher AJ, Essig S, Chin JW, Halambage UD, Aiken C, James LC.** 2014. Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly. PLoS Path. **10**(10):e1004459

341. **Bhattacharya A, Alam SL, Fricke T, Zadrozny K, Sedzicki J, Taylor AB, Demeler B, Pornillos O, Ganser-Pornillos BK, Diaz-Griffero F, Ivanov DN, Yeager M.** 2014. Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. Proc. Natl. Acad. Sci. USA **111**(52):18625-18630

342. **Price AJ, Fletcher AJ, Schaller T, Elliott T, Lee K, KewalRamani VN, Chin JW, Towers GJ, James LC.** 2012. CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. PLoS Path. **8**(8):e1002896

343. **Lee K, Mulky A, Yuen W, Martin TD, Meyerson NR, Choi L, Yu H, Sawyer SL, Kewalramani VN.** 2012. HIV-1 capsid-targeting domain of cleavage and polyadenylation specificity factor 6. J. Virol. **86**(7):3851-3860

344. **Hori T, Takeuchi H, Saito H, Sakuma R, Inagaki Y, Yamaoka S.** 2013. A carboxy-terminally truncated human CPSF6 lacking residues encoded by exon 6 inhibits HIV-1 cDNA synthesis and promotes capsid disassembly. J. Virol. **87**(13):7726-7736

345. **Fricke T, Valle-Casuso JC, White TE, Brandariz-Nunez A, Bosche WJ, Reszka N, Gorelick R, Diaz-Griffero F.** 2013. The ability of TNPO3-depleted cells to inhibit HIV-1 infection requires CPSF6. Retrovirology **10**(46) [Online] doi: 10.1186/1742-4690-10-46

346. **Mateu MG.** 2009. The capsid protein of human immunodeficiency virus: intersubunit interactions during virus assembly. FEBS J. **276**(21):6098-6109

347. **Campbell S, Vogt VM.** 1995. Self-assembly in vitro of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. J. Virol. **69**(10):6487-6497

348. **Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI.** 1999. Assembly and analysis of conical models for the HIV-1 core. Science **283**(5398):80-83

349. **Layne SP, Merges MJ, Dembo M, Spouge JL, Conley SR, Moore JP, Raina JL, Renz H, Gelderblom HR, Nara PL.** 1992. Factors underlying spontaneous inactivation and susceptibility to neutralization of human immunodeficiency virus. Virology **189**(2):695-714

350. **Bourinbaiar AS.** 1994. The ratio of defective HIV-1 particles to replication-competent infectious virions. Acta Virol. **38**(1):59-61

351. **Maertens GN, Cook NJ, Wang W, Hare S, Gupta SS, Oztop I, Lee K, Pye VE, Cosnefroy O, Snijders AP, Kewalramani VN, Fassati A, Engelman A, Cherepanov P.** 2014. Structural basis for nuclear import of splicing factors by human Transportin 3. Proc. Natl. Acad. Sci. USA **111**(7):2728-2733

352. **Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D, Bickmore W, Poeschla E, Bushman FD.** 2007. Role of PSIP1/LEDGF/p75 in Lentiviral infectivity and integration targeting. PLoS One **2**(12):e1340

353. **Krull S, Dorries J, Boysen B, Reidenbach S, Magnius L, Norder H, Thyberg J, Cordes VC.** 2010. Protein Tpr is required for establishing nuclear pore-associated zones of heterochromatin exclusion. EMBO J. **29**(10):1659-1673

354. **Cooper A, Garcia M, Petrovas C, Yamamoto T, Koup RA, Nabel GJ.** 2013. HIV-1 causes CD4 cell death through DNA-dependent protein kinase during viral integration. Nature **498**(7454):376-379

355. **Rougemaille M, Dieppois G, Kisseleva-Romanova E, Gudipati RK, Lemoine S, Blugeon C, Boulay J, Jensen TH, Stutz F, Devaux F, Libri D.** 2008. THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association. Cell **135**(2):308-321

356. **Kurshakova MM, Krasnov AN, Kopytova DV, Shidlovskii YV, Nikolenko JV, Nabirochkina EN, Spehner D, Schultz P, Tora L, Georgieva SG.** 2007. SAGA and a novel Drosophila export complex anchor efficient transcription and mRNA export to NPC. EMBO J. **26**(24):4956-4965

357. **Raices M, D'Angelo MA.** 2012. Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions. Nat. Rev. Mol. Cell Biol. **13**(11):687-699

358. **Ambrose Z, Lee K, Ndjomou J, Xu H, Oztop I, Matous J, Takemura T, Unutmaz D, Engelman A, Hughes SH, KewalRamani VN.** 2012. Human immunodeficiency virus type 1 capsid mutation N74D alters cyclophilin A dependence and impairs macrophage infection. J. Virol. **86**(8):4708-4714

359. **Mankan AK, Schmidt T, Chauhan D, Goldeck M, Honing K, Gaidt M, Kubarenko AV, Andreeva L, Hopfner KP, Hornung V.** 2014. Cytosolic RNA:DNA hybrids activate the cGAS-STING axis. EMBO J. **33**(24):2937-2946

360. **Sun L, Wu J, Du F, Chen X, Chen ZJ.** 2013. Cyclic GMP-AMP Synthase Is a Cytosolic DNA Sensor That Activates the Type I Interferon Pathway. Science **339**(6121):786-791

361. **Wu J, Sun L, Chen X, Du F, Shi H, Chen C, Chen ZJ.** 2013. Cyclic GMP-AMP Is an Endogenous Second Messenger in Innate Immune Signaling by Cytosolic DNA. Science **339**(6121):826-830

362. **Gao D, Wu J, Wu YT, Du F, Aroh C, Yan N, Sun L, Chen ZJ.** 2013. Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. Science **341**(6148):903-906

363. **Gao P, Ascano M, Zillinger T, Wang W, Dai P, Serganov AA, Gaffney BL, Shuman S, Jones RA, Deng L, Hartmann G, Barchet W, Tuschl T, Patel DJ.** 2013. Structure-Function Analysis of STING Activation by c[G(2′,5′)pA(3′,5′)p] and Targeting by Antiviral DMXAA. Cell **154**(4):748-762

364. **Diner EJ, Burdette DL, Wilson SC, Monroe KM, Kellenberger CA, Hyodo M, Hayakawa Y, Hammond MC, Vance RE.** 2013. The Innate Immune DNA Sensor cGAS Produces a Noncanonical Cyclic Dinucleotide that Activates Human STING. Cell Reports **3**(5):1355-1361

365. **Zhang X, Shi H, Wu J, Zhang X, Sun L, Chen C, Chen ZJ.** 2013. Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. Mol. Cell **51**(2):226-235

366. **Gromme M, Neefjes J.** 2002. Antigen degradation or presentation by MHC class I molecules via classical and non-classical pathways. Mol. Immunol. **39**(3-4):181-202

367. **Cresswell P, Ackerman AL, Giodini A, Peaper DR, Wearsch PA.** 2005. Mechanisms of MHC class I-restricted antigen processing and cross-presentation. Immunol. Rev. **207**:145-157

368. **Madden DR, Gorga JC, Strominger JL, Wiley DC.** 1992. The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC. Cell **70**(6):1035-1048

369. **Gao GF, Tormo J, Gerth UC, Wyer JR, McMichael AJ, Stuart DI, Bell JI, Jones EY, Jakobsen BK.** 1997. Crystal structure of the complex between human CD8alpha(alpha) and HLA-A2. Nature **387**(6633):630-634

370. **Ding YH, Smith KJ, Garboczi DN, Utz U, Biddison WE, Wiley DC.** 1998. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. Immunity **8**(4):403-411

371. **Goodsell D.** 2005. T-Cell Receptor. RCSB. [Online] URL:http://www.rcsb.org/pdb/101/motm.do?momID=63

372. **Barry M, Bleackley RC.** 2002. Cytotoxic T lymphocytes: all roads lead to death. Nat. Rev. Immunol. **2**(6):401-409

373. **Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, Eichbaum Q, Block BL, Baker B, Kadie C, Markowitz M, Jessen H, Kelleher AD, Rosenberg E, Kaldor J, Yuki Y, Carrington M, Allen TM, Mallal S, Altfeld M, Heckerman D, Walker BD.** 2008. Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. J. Virol. **82**(18):9216-9227

374. **Walker B, McMichael A**. 2012. The T-cell response to HIV. Cold Spring Harb. Perspect. Med. [Online] doi: 10.1101/cshperspect.a007054

375. **Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A, Goldstein DB.** 2007. A whole-genome association study of major determinants for host control of HIV-1. Science **317**(5840):944-947

376. **International HIV Controllers Study, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, Ripke S, Brumme CJ, Pulit SL, Carrington M, Kadie CM, Carlson JM, Heckerman D, Graham RR, Plenge RM, Deeks SG, Gianniny L, Crawford G, Sullivan J, Gonzalez E, Davies L, Camargo A, Moore JM, Beattie N, Gupta S, Crenshaw A, Burtt NP, Guiducci C, Gupta N, Gao X, Qi Y, Yuki Y, Piechocka-Trocha A, Cutrell E, Rosenberg R, Moss KL, Lemay P, O'Leary J, Schaefer T, Verma P, Toth I, Block B, Baker B, Rothchild A, Lian J, Proudfoot J, Alvino DM, Vine S, Addo MM, Allen TM, Altfeld M, Henn MR, Le Gall S, Streeck H, Haas DW, Kuritzkes DR, Robbins GK, Shafer RW, Gulick RM, Shikuma CM, Haubrich R, Riddler S, Sax PE, Daar ES, Ribaudo HJ, Agan B, Agarwal S, Ahern RL, Allen BL, Altidor S, Altschuler EL, Ambardar S, Anastos K, Anderson B, Anderson V, Andrady U, Antoniskis D, Bangsberg D, Barbaro D, Barrie W, Bartczak J, Barton S, Basden P, Basgoz N, Bazner S, Bellos NC, Benson AM, Berger J, Bernard NF, Bernard AM, Birch C, Bodner SJ, Bolan RK, Boudreaux**

ET, Bradley M, Braun JF, Brndjar JE, Brown SJ, Brown K, Brown ST, Burack J, Bush LM, Cafaro V, Campbell O, Campbell J, Carlson RH, Carmichael JK, Casey KK, Cavacuiti C, Celestin G, Chambers ST, Chez N, Chirch LM, Cimoch PJ, Cohen D, Cohn LE, Conway B, Cooper DA, Cornelson B, Cox DT, Cristofano MV, Cuchural G, Jr., Czartoski JL, Dahman JM, Daly JS, Davis BT, Davis K, Davod SM, DeJesus E, Dietz CA, Dunham E, Dunn ME, Ellerin TB, Eron JJ, Fangman JJ, Farel CE, Ferlazzo H, Fidler S, Fleenor-Ford A, Frankel R, Freedberg KA, French NK, Fuchs JD, Fuller JD, Gaberman J, Gallant JE, Gandhi RT, Garcia E, Garmon D, Gathe JC, Jr., Gaultier CR, Gebre W, Gilman FD, Gilson I, Goepfert PA, Gottlieb MS, Goulston C, Groger RK, Gurley TD, Haber S, Hardwicke R, Hardy WD, Harrigan PR, Hawkins TN, Heath S, Hecht FM, Henry WK, Hladek M, Hoffman RP, Horton JM, Hsu RK, Huhn GD, Hunt P, Hupert MJ, Illeman ML, Jaeger H, Jellinger RM, John M, Johnson JA, Johnson KL, Johnson H, Johnson K, Joly J, Jordan WC, Kauffman CA, Khanlou H, Killian RK, Kim AY, Kim DD, Kinder CA, Kirchner JT, Kogelman L, Kojic EM, Korthuis PT, Kurisu W, Kwon DS, LaMar M, Lampiris H, Lanzafame M, Lederman MM, Lee DM, Lee JM, Lee MJ, Lee ET, Lemoine J, Levy JA, Llibre JM, Liguori MA, Little SJ, Liu AY, Lopez AJ, Loutfy MR, Loy D, Mohammed DY, Man A, Mansour MK, Marconi VC, Markowitz M, Marques R, Martin JN, Martin HL, Jr., Mayer KH, McElrath MJ, McGhee TA, McGovern BH, McGowan K, McIntyre D, McLeod GX, Menezes P, Mesa G, Metroka CE, Meyer-Olson D, Miller AO, Montgomery K, Mounzer KC, Nagami EH, Nagin I, Nahass RG, Nelson MO, Nielsen C, Norene DL, O'Connor DH, Ojikutu BO, Okulicz J, Oladehin OO, Oldfield EC, 3rd, Olender SA, Ostrowski M, Owen WF, Jr., Pae E, Parsonnet J, Pavlatos AM, Perlmutter AM, Pierce MN, Pincus JM, Pisani L, Price LJ, Proia L, Prokesch RC, Pujet HC, Ramgopal M, Rathod A, Rausch M, Ravishankar J, Rhame FS, Richards CS, Richman DD, Rodes B, Rodriguez M, Rose RC, 3rd, Rosenberg ES, Rosenthal D, Ross PE, Rubin DS, Rumbaugh E, Saenz L, Salvaggio MR, Sanchez WC, Sanjana VM, Santiago S, Schmidt W, Schuitemaker H, Sestak PM, Shalit P, Shay W, Shirvani VN, Silebi VI, Sizemore JM, Jr., Skolnik PR, Sokol-Anderson M, Sosman JM, Stabile P, Stapleton JT, Starrett S, Stein F, Stellbrink HJ, Sterman FL, Stone VE, Stone DR, Tambussi G, Taplitz RA, Tedaldi EM, Telenti A, Theisen W, Torres R, Tosiello L, Tremblay C, Tribble MA, Trinh PD, Tsao A, Ueda P, Vaccaro A, Valadas E, Vanig TJ, Vecino I, Vega VM, Veikley W, Wade BH, Walworth C, Wanidworanun C, Ward DJ, Warner DA, Weber RD, Webster D, Weis S, Wheeler DA, White DJ, Wilkins E, Winston A, Wlodaver CG, van't Wout A, Wright DP, Yang OO, Yurdin DL, Zabukovic BW, Zachary KC, Zeeman B, Zhao M. 2010. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science **330**(6010):1551-1557

377. **Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, Racz P, Tenner-Racz K, Dalesandro M, Scallon BJ, Ghrayeb J, Forman MA, Montefiori DC, Rieber EP, Letvin NL, Reimann KA.** 1999. Control

of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. Science **283**(5403):857-860

378. **Schwartz O, Marechal V, Le Gall S, Lemonnier F, Heard JM.** 1996. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. Nat. Med. **2**(3):338-342

379. **Collins KL, Chen BK, Kalams SA, Walker BD, Baltimore D.** 1998. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. Nature **391**(6665):397-401

380. **Deng K, Pertea M, Rongvaux A, Wang L, Durand CM, Ghiaur G, Lai J, McHugh HL, Hao H, Zhang H, Margolick JB, Gurer C, Murphy AJ, Valenzuela DM, Yancopoulos GD, Deeks SG, Strowig T, Kumar P, Siliciano JD, Salzberg SL, Flavell RA, Shan L, Siliciano RF.** 2015. Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. Nature **517**(7534):381-385

381. **Goulder PJ, Walker BD.** 2012. HIV and HLA class I: an evolving relationship. Immunity **37**(3):426-440

382. **Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC.** 1987. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature **329**(6139):512-518

383. **Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, Keele BF, Learn GH, Turnbull EL, Salazar MG, Weinhold KJ, Moore S, B CCC, Letvin N, Haynes BF, Cohen MS, Hraber P, Bhattacharya T, Borrow P, Perelson AS, Hahn BH, Shaw GM, Korber BT, McMichael AJ.** 2009. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J. Exp. Med. **206**(6):1253-1272

384. **Carlson JM, Brumme ZL.** 2008. HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective. Microbes Infect. **10**(5):455-461

385. **Kosmrlj A, Read EL, Qi Y, Allen TM, Altfeld M, Deeks SG, Pereyra F, Carrington M, Walker BD, Chakraborty AK.** 2010. Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. Nature **465**(7296):350-354

386. **Zuniga R, Lucchetti A, Galvan P, Sanchez S, Sanchez C, Hernandez A, Sanchez H, Frahm N, Linde CH, Hewitt HS, Hildebrand W, Altfeld M, Allen TM, Walker BD, Korber BT, Leitner T, Sanchez J, Brander C.** 2006. Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. J. Virol. **80**(6):3122-3125

387. **Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, Reddy S, de Pierres C, Mncube Z, Mkhwanazi N, Bishop K, van der Stok M, Nair K, Khan N, Crawford H, Payne R, Leslie A, Prado J, Prendergast A, Frater J, McCarthy N, Brander C, Learn GH, Nickle D, Rousseau C, Coovadia H, Mullins JI, Heckerman D, Walker BD, Goulder P.** 2007. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. Nat. Med. **13**(1):46-53

388. **Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, Goepfert PA.** 2002. Magnitude of functional CD8+ T-cell responses to the gag protein

of human immunodeficiency virus type 1 correlates inversely with viral load in plasma. J. Virol. **76**(5):2298-2305

389. **Sacha JB, Chung C, Rakasz EG, Spencer SP, Jonas AK, Bean AT, Lee W, Burwitz BJ, Stephany JJ, Loffredo JT, Allison DB, Adnan S, Hoji A, Wilson NA, Friedrich TC, Lifson JD, Yang OO, Watkins DI.** 2007. Gag-specific CD8+ T lymphocytes recognize infected cells before AIDS-virus integration and viral protein expression. J. Immunol. **178**(5):2746-2754

390. **Miura T, Brockman MA, Brumme ZL, Brumme CJ, Pereyra F, Trocha A, Block BL, Schneidewind A, Allen TM, Heckerman D, Walker BD.** 2009. HLA-associated alterations in replication capacity of chimeric NL4-3 viruses carrying gag-protease from elite controllers of human immunodeficiency virus type 1. J. Virol. **83**(1):140-149

391. **Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, Chetty S, Thobakgale C, Honeyborne I, Crawford H, Matthews P, Pillay T, Rousseau C, Mullins JI, Brander C, Walker BD, Stuart DI, Kiepiela P, Goulder P.** 2006. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. J. Virol. **80**(7):3617-3623

392. **Navis M, Schellens I, van Baarle D, Borghans J, van Swieten P, Miedema F, Kootstra N, Schuitemaker H.** 2007. Viral replication capacity as a correlate of HLA B57/B5801-associated nonprogressive HIV-1 infection. J. Immunol. **179**(5):3133-3143

393. **Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ, Goedert JJ, Winkler C, O'Brien SJ, Rinaldo C, Detels R, Blattner W, Phair J, Erlich H, Mann DL.** 1996. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat. Med. **2**(4):405-411

394. **Altfeld M, Addo MM, Rosenberg ES, Hecht FM, Lee PK, Vogel M, Yu XG, Draenert R, Johnston MN, Strick D, Allen TM, Feeney ME, Kahn JO, Sekaly RP, Levy JA, Rockstroh JK, Goulder PJ, Walker BD.** 2003. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. AIDS **17**(18):2581-2591

395. **Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, Tang Y, Holmes EC, Allen T, Prado JG, Altfeld M, Brander C, Dixon C, Ramduth D, Jeena P, Thomas SA, John AS, Roach TA, Kupfer B, Luzzi G, Edwards A, Taylor G, Lyall H, Tudor-Williams G, Novelli V, Martinez-Picado J, Kiepiela P, Walker BD, Goulder PJR.** 2004. HIV evolution: CTL escape mutation and reversion after transmission. Nat. Med. **10**(3):282-289

396. **Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA, Giangrande P, Luzzi G, Morgan B, Edwards A, McMichael AJ, Rowland-Jones S.** 1997. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat. Med. **3**(2):212-217

397. **Kelleher AD, Long C, Holmes EC, Allen RL, Wilson J, Conlon C, Workman C, Shaunak S, Olson K, Goulder P, Brander C, Ogg G, Sullivan JS, Dyer W, Jones I, McMichael AJ, Rowland-Jones S, Phillips RE.** 2001. Clustered mutations in HIV-1 gag are consistently required for

escape from HLA-B27-restricted cytotoxic T lymphocyte responses. J. Exp. Med. **193**(3):375-386

398. **Goulder PJ, Brander C, Tang Y, Tremblay C, Colbert RA, Addo MM, Rosenberg ES, Nguyen T, Allen R, Trocha A, Altfeld M, He S, Bunce M, Funkhouser R, Pelton SI, Burchett SK, McIntosh K, Korber BT, Walker BD.** 2001. Evolution and transmission of stable CTL escape mutations in HIV infection. Nature **412**(6844):334-338

399. **Lichterfeld M, Kavanagh DG, Williams KL, Moza B, Mui SK, Miura T, Sivamurthy R, Allgaier R, Pereyra F, Trocha A, Feeney M, Gandhi RT, Rosenberg ES, Altfeld M, Allen TM, Allen R, Walker BD, Sundberg EJ, Yu XG.** 2007. A viral CTL escape mutation leading to immunoglobulin-like transcript 4-mediated functional inhibition of myelomonocytic cells. J. Exp. Med. **204**(12):2813-2824

400. **Schneidewind A, Brockman MA, Sidney J, Wang YE, Chen H, Suscovich TJ, Li B, Adam RI, Allgaier RL, Mothe BR, Kuntzen T, Oniangue-Ndza C, Trocha A, Yu XG, Brander C, Sette A, Walker BD, Allen TM.** 2008. Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid. J. Virol. **82**(11):5594-5605

401. **Nietfield W, Bauer M, Fevrier M, Maier R, Holzwarth B, Frank R, Maier B, Riviere Y, Meyerhans A.** 1995. Sequence constraints and recognition by CTL of an HLA-B27-restricted HIV-1 gag epitope. J. Immunol. **154**(5):2189-2197

402. **Aberham C, Weber S, Phares W.** 1996. Spontaneous mutations in the human immunodeficiency virus type 1 gag gene that affect viral replication in the presence of cyclosporins. J. Virol. **70**(6):3536-3544

403. **Qi M, Yang R, Aiken C.** 2008. Cyclophilin A-dependent restriction of Human Immunodeficiency Virus type 1 capsid mutants for infection of nondividing cells. J. Virol. **82**(24):12001-12008

404. **Tang S, Poulin L, Levy JA.** 1992. Lack of human immunodeficiency virus type 1 (HIV-1) replication and accumulation of viral DNA in HIV-1-infected T cells blocked in cell replication. J. Gen. Virol. **73**(4):933-939

405. **Goldstone DC, Walker PA, Calder LJ, Coombs PJ, Kirkpatrick J, Ball NJ, Hilditch L, Yap MW, Rosenthal PB, Stoye JP, Taylor IA.** 2014. Structural studies of postentry restriction factors reveal antiparallel dimers that enable avid binding to the HIV-1 capsid lattice. Proc. Natl. Acad. Sci. USA **111**(26):9609-9614

406. **Schneidewind A, Brumme ZL, Brumme CJ, Power KA, Reyor LL, O'Sullivan K, Gladden A, Hempel U, Kuntzen T, Wang YE, Oniangue-Ndza C, Jessen H, Markowitz M, Rosenberg ES, Sekaly RP, Kelleher AD, Walker BD, Allen TM.** 2009. Transmission and long-term stability of compensated CD8 escape mutations. J. Virol. **83**(8):3993-3997

407. **Li G, Verheyen J, Rhee SY, Voet A, Vandamme AM, Theys K**. 2013. Functional conservation of HIV-1 Gag: implications for rational drug design. Retrovirology **10**(126) [Online] doi: 10.1186/1742-4690-10-126

408. **Margot NA, Gibbs CS, Miller MD.** 2010. Phenotypic susceptibility to bevirimat in isolates from HIV-1-infected patients without prior exposure to bevirimat. Antimicrob. Agents Chemother. **54**(6):2345-2353

409. **Neira JL.** 2009. The capsid protein of human immunodeficiency virus: designing inhibitors of capsid assembly. FEBS J. **276**(21):6110-6117

410. **Cao J, Isaacson J, Patick AK, Blair WS.** 2005. High-throughput human immunodeficiency virus type 1 (HIV-1) full replication assay that includes HIV-1 Vif as an antiviral target. Antimicrob. Agents Chemother. **49**(9):3833-3841

411. **Blair WS, Cao J, Fok-Seang J, Griffin P, Isaacson J, Jackson RL, Murray E, Patick AK, Peng Q, Perros M, Pickford C, Wu H, Butler SL.** 2009. New small-molecule inhibitor class targeting human immunodeficiency virus type 1 virion maturation. Antimicrob. Agents Chemother. **53**(12):5080-5087

412. **Shi J, Zhou J, Halambage UD, Shah VB, Burse MJ, Wu H, Blair WS, Butler SL, Aiken C.** 2015. Compensatory Substitutions in the HIV-1 Capsid Reduce the Fitness Cost Associated with Resistance to a Capsid-Targeting Small-Molecule Inhibitor. J. Virol. **89**(1):208-219

413. **Fricke T, Buffone C, Opp S, Valle-Casuso J, Diaz-Griffero F**. 2014. BI-2 destabilizes HIV-1 cores during infection and prevents binding of CPSF6 to the HIV-1 capsid. Retrovirology **11**(120) [Online] doi: 10.1186/s12977-014-0120-x