Probing the expanded architecture of the lentiviral intasome

Vidya Chivukula

University College London

and

The Francis Crick Institute

PhD Supervisor: Prof. Peter Cherepanov

A thesis submitted for the degree of

Doctor of Philosophy

April 2021

Declaration

I, Vidya Chivukula, 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.

Abstract

Integration of viral DNA into a host chromosome, by action of the viral enzyme integrase (IN), is an essential step of the retroviral lifecycle. To fulfil its function, IN assembles into a multimer on the viral DNA ends, forming a highly stable nucleoprotein complex known as the intasome. The intasome architecture varies between the retroviral genera, and the maedi-visna virus (MVV) intasome contains a homo-hexadecamer (tetramer-of-tetramers) of IN. The conserved intasome core, observed in all structurally characterized retroviral intasomes, is formed between a pair of MVV IN tetramers, each providing one active site, and is completed by the insertions of the synaptic C-terminal domains (CTDs) donated by a pair of flanking IN tetramers. It was argued that this configuration is necessitated by the propensity of lentiviral INs to form tetramers in solution and the α-helical structure of the linkers connecting the catalytic core domain (CCD) and the CTD. Within the MVV IN hexadecamer, a pair of CTD tetrads bridge the IN tetramers by forming intra- and inter-tetramer interactions. Using site directed mutagenesis, the importance of these distinctive structural features was probed. The mutations disrupting the CTD-CTD interfaces or destabilizing the α-helical configuration of the CCD-CTD linkers perturb the ability of MVV IN to form multimers, assemble into stable intasomes and strongly affect its strand transfer activity *in vitro*. Moreover, these mutations strongly compromised infectivity of single-cycle MVV vector in cells.

Lentiviral integration distinctively favours actively transcribing genes, which is facilitated by the interaction of IN with LEDGF/p75, a chromatin-bound adaptor protein. The presence of LEDGF/p75 was essential to observe MVV IN strand transfer activity or intasome assembly *in vitro*. To determine the importance of LEDGF/p75 for integration in the context of viral integration in cells, infectivity of MVV-derived vector in LEDGF-knockout cells was tested. Although ablation of the host factor in human and sheep cells did not lead to a reproducible reduction of infectivity, it led to a notable shift in integration pattern, away from the usual gene bodies and transcription units.

Collectively, these observations indicate that the hexadecameric architecture is critical for the MVV IN function and suggest that LEDGF/p75 is relevant in targeting its integration towards favored regions of host genome.

Impact Statement

Retroviral integration is an essential step in the viral lifecycle, which is catalysed by the viral enzyme integrase (IN). Integration is carried out by a multimer of IN subunits assembled at the viral DNA (vDNA) ends, forming an intasome. As such, integration is an irreversible step, wherein vDNA is permanently inserted into the target genome. Lentiviruses, including HIV-1 were shown to assemble into the largest known intasome complexes *in vitro*, to perform this crucial step.

Academic impact

The research presented in this thesis has successfully determined that the unique, extended intasome architecture observed in lentiviruses is functionally relevant for both catalysing the integration reactions *in vitro*, as well as for the virus to establish infectivity in cells. These findings were presented at the Frontiers in Retrovirology and Cold Spring Harbour Laboratory Retrovirology (virtual) meetings, and a manuscript is currently in preparation for publication. As a result of this research, techniques were developed or optimised for use in the Cherepanov lab, which can be useful for future studies. These include, visualisation of MVV vectors by thin-section EM and quantification of viral titre by product enhanced reverse transcriptase (PERT) assay, which are in use by collaborators and other labs at the Francis Crick Institute. Furthermore, clonal and polyclonal sheep cells knocked out for the expression of LEDGF/p75 and HRP2 were generated using CRISPR/Cas9 based methods and a plasmid encoding MVV LRT product was also generated, which will be available for future studies.

Medical/Societal Impact

The research within this thesis was aimed to understand the functional consequences of the extended lentiviral intasome architecture. This can help further the understanding of why lentiviruses assume a unique, rather complex intasome assembly. High-resolution lentiviral intasome structures form HIV-1, MVV and SIV_{rcm} are of great importance in designing drug molecules for inhibiting the integration reaction and subsequently use them in the clinical context.

Acknowledgement

I would like thank Prof. Peter Cherepanov for giving me the opportunity of pursuing this PhD and training me to become a scientist. I would also like to thank my thesis committee, Dr. Jonathan Stoye, Prof. Ariberto Fassati and Dr. Katrin Rittinger for their input and feedback on my project. A special thanks to Ari, for providing me with guidance when I needed it. A big thanks to the collaborators of this project, especially the members of the Engelman lab and Alan himself, who always patiently provided scientific insight in this project. I would like to also thank Dr. Raffaella Carzaniga from the EM core facility at the Francis Crick Institute, who taught me everything I know about thin-section EM and patiently optimised the protocols with me. Also the various other core facilities at the Crick, including Vicky Deering from the Genomics equipment park, the entire cell-services and flow cytometry teams who have continually provided me with resources and technical support. The Bishop lab at the Crick, who were kind enough to allow me to use their tissue culture room.

I would also like to thank members of the Cherepanov lab members, in particular Allison, Val, Annachiara, Nic and Anne, who have not only contributed to my scientific journey, but also provided me with emotional support. Chloe, from the Bishop lab who has suffered through endless discussions with me during our times together in the tissue culture. A special mention to my fellow PhD students, Rowan, Souradeep and Lavanya, who have been through thick and thin with me in the course of this PhD.

Most of all, I would like to thank my family for their constant love and support. My husband, Mithun who not only exposed me to the finer things in life, but is also my one constant pillar of support. My mother Rajeswari, who has been a source of inspiration by being overcoming adversities, and showing me how to be a productive member of the society. My brother Chinmayanand, for keeping me grounded and being the best sibling one can ask for. My mother in law, Rukmaniben, who ensured that I never missed home. Both my grandmothers, who are extremely strong women and constantly pushed me to achieve academic success. Finally, my late father, who I lost during the final year of my PhD. He was extremely proud of me doing a PhD and although he was not a scientist, he would attempt to read all my scientific reports.

‘Daddy, I wish you were here to read my thesis!’

Table of Contents

Table of figures

List of tables

Abbreviations

AIDS	acquired immunodeficiency syndrome
ALLINI	allosteric integrase inhibitor
ANXA6	annexin A6
APOBEC3G	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G
APS	ammonium persulphate
ART	effective antiretroviral therapy
ARV	AIDS-associated retrovirus
ASLV	avian sarcoma leukosis virus
ATL	adult T-cell leukaemia
AZT	azidothymidine
BAF	barrier-to-autointegration factor
BCA	bicinchoninic acid
BET	bromodomain and extra-terminal domain
BIV	bovine immunodeficiency virus
BLV	bovine leukaemia virus
bp	base pair(s)
BSA	bovine serum albumin
BTP	bistris-propane
CA	capsid
CAGP	chicken beta-actin/CMV promoter/enhancer element
CCD	catalytic core domain
CEAV	caprine arthritis-encephalitis virus
CFIm	cleavage factor Im
CHAPS	3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate

CIC	conserved intasome core
CLIP-Seq	cross-linking immunoprecipitation combined with high-throughput sequencing
CMVP	cytomegalovirus immediate early promoter
cPPT/cts	central polypurine tract and central termination sequence
CPSF6	cleavage and polyadenylation specificity factor 6
CPT-3	sheep choroid plexus cells, immortalised with TERT
crRNA	crispr RNA
Cryo-EM	cryo-electron microscopy
CSK I/II	cytoskeletal I/II buffer
CTD	C-terminal domain
CTE	constitutive transport element
CypA	cyclophilin A
d	day(s)
dA	adenine
dC	cytosine
ddC	2′,3′-dideoxycytidine
DDSA	dodecenylsuccinic anhydride
dG	guanine
DMEM	Dulbecco’s modified Eagle’s medium
DMP-30	tris (dimethylaminomethyl)phenol
DNA	deoxyribo-nucleic acid
dNTP	deoxynucleotide triphosphate
DSBs	double stranded breaks
dsDNA	double-stranded DNA
dT	thymine

DTG	dolutegravir
DTT	dithiothreitol
EIAV	equine infectious anaemia virus
EM	electron microscopy
ESCRT	endosomal sorting complex required for transport
FACS	fluorescence activated cell sorting
FACT	facilitates chromatin transcription
FBS	foetal bovine serum
FIV	feline immunodeficiency virus
FS	full-site (strand transfer product)
FSC-A	area plot of forward scatter
FSC-H	height plot of forward scatter
GFP	green fluorescence protein
gRNA	guide RNA
h	hour(s)
HAM/TSP	HTLV-I associated myelopathy/tropical spastic paraparesis
HDGF	hepatoma derived-growth factor
HEK-293T	human embryonic kidney 293 cells expressing a mutant version of the SV40 large T antigen
His_6	hexahistidine tag
HIV-1/-2	human immunodeficiency virus type 1 or 2
HKO	HRP2 knockout cells
HMG1A	high-mobility group protein 1A
HRP	horseradish peroxidase
HRP	HDGF-related protein
HRV	human rhinovirus

HS	half-site (strand transfer product)
HTLV-I/-II	human T-lymphotropic virus type I or II
IBD	integrase-binding domain
IN	integrase
INSTI	integrase strand transfer inhibitor
IP6	inositol hexaphosphate
IPTG	isopropyl β-d-1-thiogalactopyranoside
kb	kilo base(s)
LADs	lamina associated domains
LAR	luciferase assay reagent
LAV	lymphadenopathy-associated virus
LB broth	Luria-Bertani broth
LEDGF/p75	lens-epithelium derived growth factor splice variant p75
LHKO	LEDGF/p75 and HRP2 knockout cells
LKO	LEDGF/p75 knockout cells
LRT product	late-reverse transcription product
LTR	long terminal repeat
MA	matrix
MCM10	minichromosome maintenance complex component 10
MEF	mouse embryonic fibroblast
min	minute(s)
MLV	murine leukaemia virus
MMTV	mouse mammary tumour virus
MNA	methyl-5-norbornene-2,3-dicarboxylic anhydride
MPMV	Mason-Pfizer monkey virus
MRC	matched random control

MVV	maedi visna virus
NaBu	sodium butyrate
NC	nucleocapsid
NED	N-terminal extension domain
Nef	negative factor
NHEJ	non-homologous end joining
NLS	nuclear localisation signal
NMR	nuclear magnetic resonance
NNRTI	non-nucleoside reverse transcriptase inhibitor
NRTI	nucleoside reverse transcriptase inhibitor
NTD	N-terminal domain
OD	optical density
PBS	primer-binding site (in the context of viral genome)
PBS	phosphate buffered saline
PBS-T	phosphate buffered saline with 0.001% tween
PCR	polymerase chain reaction
PEI	polyethylenimine
Pfu	*Pyrococcus furiosus* DNA polymerase
PFV	prototype foamy virus
PIC	pre-integration complex
PLB	passive lysis buffer
PMSF	phenylmethylsulfonyl fluoride
PN	peripheral nuclear
PP2A	serine/threonine-protein phosphatase 2A
PPT	polypurine tract
PVDF	polyvinylidene difluoride

qPCR	quantitative PCR
R	short repeat (R) sequences (in the context of viral genome)
Rev	regulator of expression of virion proteins
RIGs	recurrent integration targeted genes
RLU	relative luciferase light units
RNA	ribo-nucleic acid
RNP	ribonucleoprotein
rpm	revolutions per minute
RRE	Rev-responsive element
RT	reverse transcriptase
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
sec	second(s)
SEC-MALLS	size exclusion chromatography coupled with multi-angle laser light scattering
SERINC5	serine incorporator 5
SIV	simian immunodeficiency virus
SIV_{cpzPts}	SIV from *Pan troglodytes schweinfurthii*
SIV_{cpzPtt}	SIV from *Pan troglodytes troglodytes*
SIV_{gor}	SIV from *Gorilla gorilla gorilla*
SIV_{rcm}	SIV from red-capped mangabey
SIV_{smm}	SIV from *Cerocebus atys*
SPAD	speckle associated domain
SRLV	small ruminant lentivirus
SSC-A	area plot of side scatter
STC	strand transfer complex
STLV	simian T-lymphotropic virus

TAE	tris acetate-EDTA
TALEN	transcription activator-like effector nuclease
Taq	*Thermus aquaticus* DNA polymerase
Taq-PERT	TaqMan probe-based product enhanced reverse transcriptase
Tat	trans-activator of transcription
TCC	target capture complex
TEMED	N,N,N',N'-tetramethylethane-1,2-diamine
TERT	telomerase reverse transcriptase
TGS	tris-glycine SDS buffer
TSS	transcription start site
TU	transcription unit
U2OS	human bone osteosarcoma epithelial cells
U3	unique 3′ region
U5	unique 5′ region
UTR	untranslated region
vDNA	viral DNA
Vif	viral infectivity factor
Vpr	viral protein R
Vpu	viral protein U
VSV-G	vesicular somatic virus G protein
WDSV	walleye dermal sarcoma virus
W_{Nuc}	nucleophilic water molecule
WPRE	woodchuck hepatitis virus post-transcriptional regulatory element
WT	wild type

Chapter 1. Introduction

1.1 Retroviruses

1.1.1 **Classification and Nomenclature**

Retroviruses or "reverse transcribing viruses" are characterized by conversion of their RNA genomes into double-stranded DNA (dsDNA) for integration into the host genome. As such, these species provided the first example of a reverse direction transfer of genetic information (Weiss, 1996), using a polymerase that can synthesise DNA using an RNA template (Temin and Mizutani, 1970, Baltimore, 1970). The viral family *Retroviridae* is divided into two subfamilies: *Orthoretroviruses* and *Spumaviruses*.

Orthoretroviruses are further classified into six genera (Figure 1-1), including alpharetrovirus (*e.g*. avian sarcoma leukosis virus, ASLV), betaretrovirus (*e.g*. mouse mammary tumour virus, MMTV), gammaretrovirus (*e.g*. murine leukaemia virus, MLV), deltaretrovirus (*e.g*. human T-lymphotropic viruses I and II, HTLV-I/II; simian T-lymphotropic virus, STLV; and bovine leukaemia virus, BLV), epsilonretrovirus (walleye dermal sarcoma virus, WDSV), and lentivirus (*e.g*. human immunodeficiency virus types 1 and 2, HIV-1, HIV-2; diverse simian immunodeficiency viruses, SIVs; maedi-visna virus, MVV) (Coffin et al., 1997). Lentiviruses acquired their name from causing slow developing chronic infections in their hosts. These pathogens can target lungs, joints, nervous system, haematopoietic system and the immune system (Haase, 1986). Lentiviruses are the only retroviral genus that do not require the host cells to undergo cell division to establish infection (Lewis and Emerman, 1994).

Spumaviruses (*e.g*. prototype foamy virus, PFV) are characterised by the presence of dsDNA, due to early onset of reverse transcription occurring within virions (Moebes et al., 1997, Yu et al., 1999). Unlike other retroviruses, the majority of spumaviruses are assembled by budding into the endoplasmic reticulum (Goepfert et al., 1995) and released from the host cells potentially by exocytosis (Meiering and Linial, 2001).

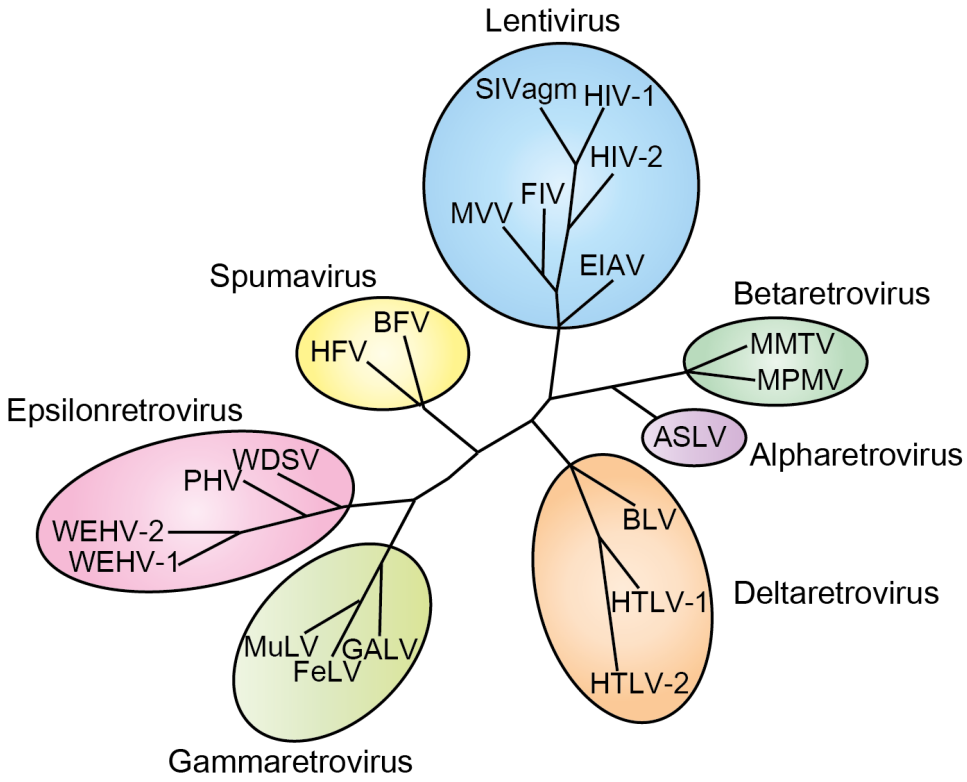

Figure 1-1 Phylogenetic tree of *Retroviridae*.

The family *Retroviridae* is divided into two sub-families: spumaviruses and orthoretroviruses. The latter subfamily contains six genera, including alpha- to epsilon-retroviruses and lentiviruses. Figure from Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses. (2012) Ed: King, A.M.Q., Adams, M.J., Carstens, E.B. & Lefkowitz, E.J. San Diego: Elsevier, under Creative Commons Attribution- ShareAlike 4.0 International License with permissions.

1.1.2 **Human and primate retroviruses**

HTLV-I and HTLV-II were isolated the from patients suffering with different types of T-cell lymphomas (Poiesz et al., 1980, Kalyanaraman et al., 1982), and determined as causative agents of adult T-cell leukaemia (ATL) and HTLV-I associated myelopathy/tropical spastic paraparesis (HAM/TSP) (Hinuma et al., 1981, Gessain et al., 1985, Osame et al., 1986). These were classified as retroviruses, due to the presence of viral RT and homology in their nucleotide sequences with the other known retroviruses (Kalyanaraman et al., 1981). The first human lentivirus was isolated as the causative agent of the acquired immunodeficiency syndrome (AIDS). It was found to bear similarities to HTLV-I and initially named HTLV-III, AIDS-associated retrovirus (ARV), or lymphadenopathy-associated virus (LAV) (Gallo et al., 1984, Popovic et al., 1984). However, it was soon shown to be genetically distant from HTLV-I/II, to possess increased cytopathic effects and to generate higher amounts of unintegrated viral DNA (vDNA) forms (Shaw et al., 1984). The species was subsequently renamed as HIV (Case, 1986, Coffin et al., 1986). HIV-2 was identified as the second additional etiologic agent of AIDS but antigenically distinct from HIV-1 (Clavel et al., 1986). This was closely related to SIV isolated from macaques (Chakrabarti et al., 1987, Guyader et al., 1987). The viral loads in patients infected with HIV-2 were lower than those observed in people infected with HIV-1 (Popper et al., 2000, Berry et al., 2002). Since their discovery, considerable resources were directed to the study of the HIV viruses. In 2020, the HIV/AIDS research budget of the US National Institutes of Health exceeded $3.5 billion. As a result of collective world-wide efforts, HIV-1 is arguably the most studied human pathogen to-date.

HIV-1 subspecies nomenclature was standardised in 2000. The virus comprises closely-related groups, M (main), N (non-M and non-O), and O (outlier) (Robertson et al., 2000), P (Plantier et al., 2009, Vallari et al., 2011). New groups were proposed to be named consecutively through the English alphabet: Q, R, *etc*. Based on their lineages, group M can further be divided into subtypes, A to D, F to H, J and K. Subtypes with distinctive lineages, yet not distinct enough for a new subtype will be designated numerically as sister clades (Robertson et al., 2000). Eight HIV-2 lineages were identified (groups A-H), from which only groups A and B were spread

throughout western Africa and Cote d'Ivoire, respectively (Damond et al., 2001, Pieniazek et al., 1999). HIV-2 from the other groups were isolated from single individuals alone (Sharp and Hahn, 2011).

Based on the available sequence information, HIV-1 and HIV-2 adapted to infecting humans following cross-species transmissions from nonhuman primates: chimpanzees, gorillas, and sooty mangabeys. Various groups of HIV-1 have originated from *Pan troglodytes troglodytes* (SIV_{cpzPtt}), *Gorilla gorilla gorilla* (SIV_{gor}), and Eastern subspecies of chimpanzees, *Pan troglodytes schweinfurthii* (SIV_{cpzPts}) (Peeters et al., 1989, Gao et al., 1999, Van Heuverswyn et al., 2006). HIV-2 originated from the SIV species prevalent among sooty mangabeys, *Cerocebus atys* (SIV_{smm}) (Gao et al., 1992).

1.1.3 **Non-primate lentiviruses**

Although the discovery and characterization of non-primate lentiviruses predates that of HIV-1, they are studied considerably less well. Bovine immunodeficiency viruses (BIVs) are found in both domestic cattle and wild buffalo (Van Der Maaten et al., 1972), equine infectious anaemia virus (EIAV) in horses (Charman et al., 1976), small ruminant lentiviruses (SRLVs) in goats and sheep (Pepin et al., 1998), and feline immunodeficiency viruses (FIVs) infect cats (Pedersen et al., 1987).

The SRLVs include MVV and caprine arthritis-encephalitis virus (CAEV) (Pepin et al., 1998). The SRLV phylogenetic relationships were established from full-length genome sequences of the MVV prototype strains (Sonigo et al., 1985, Querat et al., 1990, Sargan et al., 1991) and the CAEV type strain (Saltarelli et al., 1990). The SRLVs are classified into five related groups (A to E), based on the 1.8-kb *gag-pol* and 1.2-kb *pol* gene sequences. Group A, is further divided into eighteen subtypes, A1 to A18, which include genetically and geographically heterogenous MVVs (Shah et al., 2004, Giammarioli et al., 2011, Olech et al., 2019), while Group B is of the CAEV type and divided into four subtypes, B1 to B4 (Bertolotti et al., 2011, Olech et al., 2019). Groups C, D and E are a few isolates organised based on their geographical areas. Norwegian strains in group C (Gjerset et al., 2006), strains originating from Switzerland and Spain in group D (Reina et al., 2006) and group E

(E1 and E2) consisting strains from Northern Italy and Sardinia (Grego et al., 2007, Olech et al., 2012).

MVV is the causative agent of slow infections in sheep, especially targeting lungs and the central nervous system (Haase, 1975). Although MVV was classified as a lentivirus, the initial evidence generated in tissue culture pointed to a lack of viral integration in infected cells (Haase et al., 1982). Viral RNA synthesis was from extrachromosomal DNA and a high level of unintegrated vDNA was detected in infected cells (Harris et al., 1984). Subsequently, cells infected with MVV in tissue culture were sequenced and genomic-U3 junctions were detected only when infected with viruses carrying the wild type (WT) form of integrase (IN). This confirmed integration through the action of MVV IN, in spite of the presence of high levels of unintegrated vDNA in infected cells (List and Haase, 1997).

1.2 Epidemiology of HIV and SRLVs

The first HIV-1 group M common ancestor was dated to the early 20th century (Tebit and Arts, 2011), however AIDS was characterised as a disease in 1981. An estimated 75.7 million people have been infected with HIV since the start of the epidemic (1981 - until end of 2019), with 32.7 million deaths caused by AIDS-related illnesses and around 35 million people globally living with HIV (UNAIDS, 2020). In 2019 alone, there were 1.7 million new HIV infections and 690,000 deaths caused by AIDS-related illnesses (UNAIDS, 2020). Although the number of new infections have declined since the peak in 1998 by 40% and the total number of deaths caused by AIDS-related illnesses have reduced by 60% since its peak in 2004 (UNAIDS, 2020), partly due to the advancement of the anti-retroviral drug therapy, there still is no effective cure against HIV infection.

SRLVs undergo viral genome integration (List and Haase, 1997), however no evidence for SRLV genome in the germ line of sheep and goats was observed (Blacklaws et al., 2004). Although the transplacental route of viral transmission was relatively low (Houwers et al., 1983), viral titre was present in mammary gland secretions (colostrum or milk) from some of the infected animals (Lerondelle and Ouzrout, 1990). Lambs or kid goats fed with this milk were found seropositive in 2 to 3 months (Schipper et al., 1983, Adams et al., 1983). Two effective methods for

controlling the vertical spread of ovine progressive pneumonia were determined, wherein the seropositive sheep were culled or lambs from seropositive ewes were isolated and reared separately (Light et al., 1979, Cutlip and Lehmkuhl, 1986). Horizontal transmission of SRLVs was observed with a potential to spread through respiratory secretions (Dawson et al., 1985), along with live animal trading, by animals imported into Iceland (Sigurdsson, 1954), Canada (Dukes et al., 1979) and Hungary (Suveges et al., 1973). In one instance, infection was carried into Finland from Sweden in 1981, which led to the culling of the entire primary infection flock in 1995 (Sihvonen et al., 1999). Currently, MVV is spread worldwide, with the exceptions of Iceland, Australia and New Zealand (Kalogianni et al., 2020), although the total economic burden of this virus has not been estimated.

1.3 Retroviral genome organisation

Retroviruses package two copies of their genomic RNA, consisting of single-stranded plus-strand RNA approximately 9-kb in length and flanked at both ends with short repeat (R) sequences. The retroviral genome comprises three canonical genes: *gag*, *pol* and *env* along with species-specific set of accessory genes (Balvay et al., 2007) (Figure 1-2). During reverse transcription, long terminal repeats (LTR) are synthesised, which flank both the ends of the resulting vDNA molecule (Figure 1-2). The *gag* gene is responsible for the synthesis of the major viral polyprotein Gag, accounting for ~50% of the viral particle mass. Upon proteolytic processing, Gag yields the viral structural components, including membrane binding matrix (MA), the bipartite capsid (CA) and the nucleic acid-binding nucleocapsid (NC) proteins (Mattei et al., 2016b). HIV-1 Gag is the only viral protein required for the assembly and release of immature virus-like particles from cells (Gheysen et al., 1989, Borsetti et al., 1998, Lingappa et al., 2014). Viral enzymes, reverse transcriptase (RT), IN and protease are encoded by the *pol* gene. The envelope polyprotein is synthesised from the *env* gene. HIV-1 envelope glycoprotein precursor gp160 undergoes cleavage in host cells by furin-like proteases, giving rise to gp120 and gp41, which make up the mature envelope protein containing three copies of both subunits.

In addition to the three canonical genes, most retroviruses harbour genus- and species-specific accessory genes. Arguably, lentiviruses possess the most sophisticated and diverse array of accessory proteins, which are expressed through

extensive alternative splicing. The HIV-1 genome contains six genes that synthesise regulatory and accessory proteins (Figure 1-2), which perform a diverse range of functions to enable efficient viral replication and help the virus to evade host restriction factors. Trans-activator of transcription (Tat) plays an essential role in facilitating transcription from the viral LTR promoter (Dayton et al., 1986, Fisher et al., 1986), while the regulator of expression of virion proteins (Rev) is involved in viral RNA export from the nucleus (Sodroski et al., 1986, Emerman et al., 1989, Felber et al., 1989). The negative factor (Nef) and accessory viral protein U (Vpu) collaborate in downregulating CD4 expression surface of infected cells (Garcia and Miller, 1991, Willey et al., 1992), to prevent its interference with the subsequent viral release (Ross et al., 1999, Lama et al., 1999). Nef also plays a role in evading the innate immune response of the host cells (Schwartz et al., 1996, Stumptner-Cuvelette et al., 2001) and prevents incorporation of the host antiviral protein, serine incorporator 5 (SERINC5), into HIV-1 virions (Rosa et al., 2015). Vpu antagonises host protein tetherin, which interferes with viral particle release (Neil et al., 2008). HIV-1 viral infectivity factor (Vif) protein promotes ubiquitination of another potent restriction factor, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G (APOBEC3G) (Conticello et al., 2003, Marin et al., 2003, Stopak et al., 2003). Akin to Vif, viral protein R (Vpr) was shown to promote ubiquitination of host proteins that could interfere with the viral replication, such as uracil DNA glycosylase-2 (Ahn et al., 2010), telomerase reverse transcriptase (TERT) (Wang et al., 2013), and minichromosome maintenance complex component 10 (MCM10) (Romani et al., 2015). Vpr is also implicated in downregulating the host immune responses (Khan et al., 2020, Gibbons et al., 2020) to prevent anti-viral responses being generated by the infected cells.

MVV *pol* gene encodes for protease, RT, dUTPase and IN, while *env* encodes for gp150, which is the precursor for the mature MVV envelope glycoprotein subunits gp135 and gp44 (Vigne et al., 1982). The viral dUTPase prevents the incorporation of dUTP in vDNA during reverse transcription (Turelli et al., 1996). MVV also encodes for the regulatory *rev* gene, along with the accessory genes, including *vpr-like* (Villet et al., 2003) and *vif* (Sonigo et al., 1985, Kristbjörnsdóttir et al., 2004).

At the other end of the retroviral complexity spectrum, the genomes of ASLV and MLV encode only *gag*, *pol* and *env*, without any known additional genes, with their

genomes being approximately 7.2 kb and 8.3 kb long, respectively (King et al., 2011). Nonetheless, using an alternative start codon, MLV expresses a specialised form of Gag, (glycoGag), that aids in viral budding from lipid rafts (Nitta et al., 2010) and implicated in antagonism of the host restriction factor SERINC5 (Rosa et al., 2015).

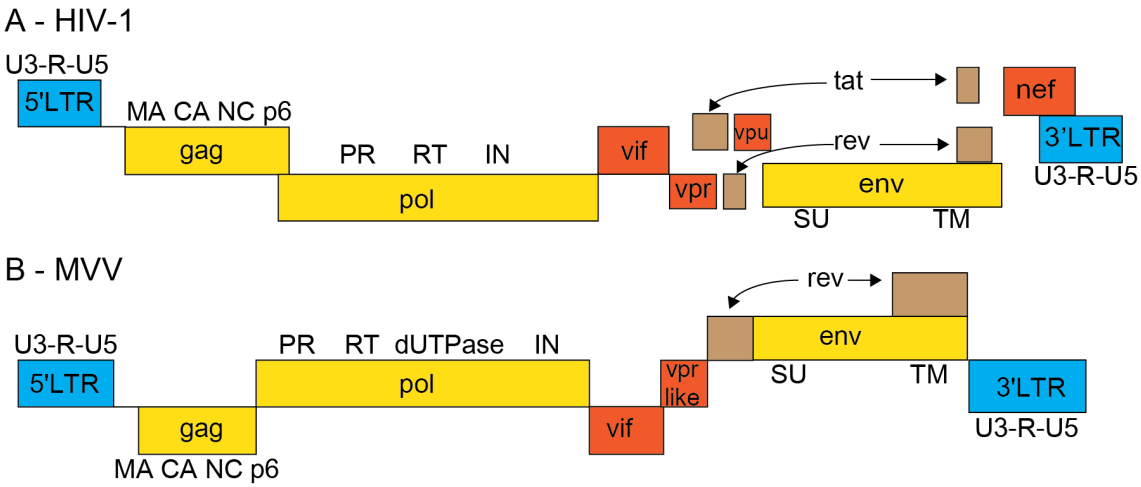

Figure 1-2 Lentiviral genome organisation.

The lentiviral genome is ~9 kb in length, harbouring three primary genes: *gag*, *pol* and *env* (yellow boxes) and flanked by two long terminal repeats (LTRs) (blue boxes). The LTRs contain unique regions (U3 and U5), separated by a repeat region (R). (A) HIV-1 *gag* encodes for matrix (MA), capsid (CA), nucleocapsid (NC) and p6 proteins. Its *pol* gene encodes the viral enzymes protease (PR), reverse transcriptase (RT) and integrase (IN), while the *env* gene is responsible for the synthesis of the glycoproteins, surface unit (SU) and transmembrane (TM) proteins. HIV-1 also contains the essential regulatory genes, *tat* and *rev* (brown), along with the accessory genes *vif*, *vpr*, *vpu* and *nef* (orange). (B) In addition to the three viral enzymes, MVV *pol* gene also encodes for dUTPase. MVV also contains the regulatory *rev* (brown), and accessory *vif* and *vpr-like* genes (orange).

1.4 Retroviral replication cycle

Retroviral lifecycle is traditionally divided into two phases – early and late. The early phase begins from the virus-cell interaction through to viral cDNA integration, while the late phase includes steps from viral gene expression to viral particle release and maturation (Nisole and Saïb, 2004) (Figure 1-3). Adsorption of viral particles to the surface of their target cells is the initial step of the viral lifecycle (Nisole and Saïb, 2004).

Viral envelope glycoproteins interact with specific cell surface protein receptors for target cell entry. Primate lentiviruses, including HIV-1, require cellular surface expression of CD4 (Maddon et al., 1986) along with a co-receptor, which in the case of HIV-1 could either be CCR5 (Deng et al., 1996) or CXCR4 (Feng et al., 1996), for entry into cells. CD4 is expressed on T-cells, monocytes and dendritic cells, as well as glial cells in the brain, which are major host cells for primate lentiviruses (Wyatt and Sodroski, 1998). Retroviral entry is a complex mechanism, which is best understood for HIV-1 (Nisole and Saïb, 2004). HIV-1 entry into $CD4^+$ cells is pH-independent and occurs through direct fusion of the viral envelope with the plasma membrane (Stein et al., 1987). Although the precise mechanism of HIV-1 entry into cells is currently unknown, it is well established that the mature envelope glycoprotein trimers, comprising of the surface protein gp120 and transmembrane gp41, undergo CD4-induced conformational changes (Trkola et al., 1996, Kwong et al., 1998). This enables the envelope protein trimers to attain an open configuration (Ozorowski et al., 2017), which allows their interaction with the co-receptors. The coreceptor binding causes further conformational changes in gp41 (Pancera et al., 2010), which subsequently leads to the fusion of viral and cellular membranes (Blumenthal et al., 2012).

Cells of monocyte or macrophage lineages are the primary targets for MVV infection (Narayan et al., 1982). A sheep mannose receptor, C-type lectin, attached to the cell membrane, was identified as a cellular receptor for MVV (Crespo et al., 2011). Although CD4 and chemokine receptors (CCR5 and CXCR4) are not essential for MVV infection, their presence helps in the formation of cell fusion or syncytium caused as a consequence of MVV infection (Hovden and Sommerfelt, 2002).

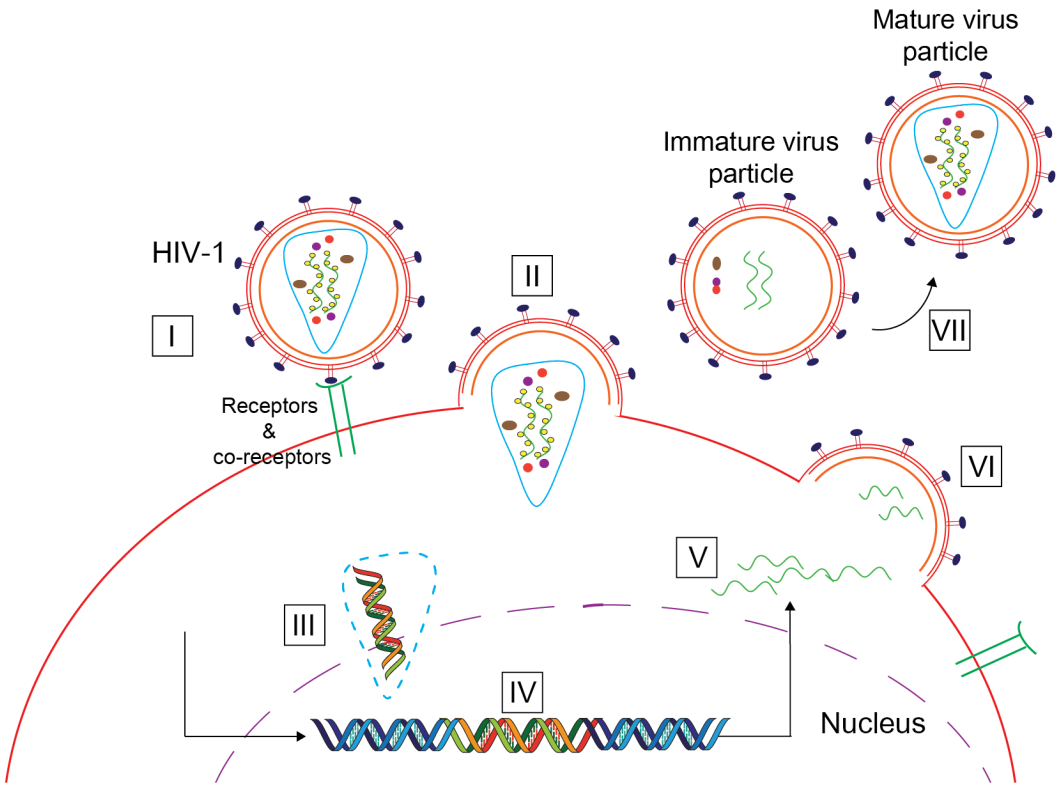

Figure 1-3 Retroviral (HIV-1) lifecycle.

Major steps of retroviral lifecycle including the early phase steps of virion attachment (I) and entry into the host cell (II), followed by reverse transcription, and uncoating (III) and integration (IV), and the late phase steps of viral gene expression (V), budding (VI) and maturation (VII) of viral particle.

Following the fusion of viral envelope and cellular membrane, the viral core is released into cytoplasm (Figure 1-3). HIV-1 core is a highly ordered conical lattice, which is composed of approximately 250 CA hexamers and 12 pentamers and encapsulates two copies of single-stranded viral genomic RNA along with its associated proteins (Zhao et al., 2013, Mattei et al., 2016a). The CA multimers comprising the viral core are stabilised by binding inositol hexaphosphate (IP6), a host cell-derived polyanionic metabolite (Dick et al., 2018, Mallery et al., 2018). Following entry into a host cell, the intact viral capsid core, and its interaction with the host factors, cyclophilin A (CypA) and the cleavage and polyadenylation specificity factor subunit 6 (CPSF6), were shown to play critical roles in protecting the reverse transcribing viral genome from recognition by host innate immunity (Schaller et al., 2011, Rasaiyaah et al., 2013, Siddiqui et al., 2019, Sumner et al., 2020).

Viral RT performs both the enzymatic activities required for reverse transcription, including DNA synthesis (DNA polymerase activity) and degradation of the viral RNA template (RNase H activity). The DNA polymerase activity of RT can copy both RNA and DNA templates, while RNase H activity cleaves RNA, only in the context of a DNA/RNA hybrid. RT uses a selectively-packaged host tRNA, encapsulated in the viral core, as a primer to initiate the synthesis of the negative strand of vDNA (Marquet et al., 1995, Isel et al., 2010). The primer, $tRNA^{Lys3}$ in case of HIV-1 (Ratner et al., 1985, Litvak et al., 1994), anneals to a complementary sequence near the 5′UTR of the viral genomic RNA, termed the primer-binding site (PBS). The R sequence allows the complementary portion of minus-strand DNA to bind to the R sequence at the 3′end of viral RNA, thereby continuing the synthesis of the negative strand. This template switching event is termed the first jump or minus-strand transfer (Peliska and Benkovic, 1992, Sarafianos et al., 2009).

A polypurine tract (PPT) directly abutting the U3 region at the 3′end of viral RNA is resistant to the RNase H activity of the RT, which allows it to act as a primer for plus-strand DNA synthesis (Luo et al., 1990). In addition, the central PPT present in the middle of the HIV-1 genome, acts as a secondary site of initiation of plus-strand DNA synthesis (Charneau et al., 1992). The plus-strand DNA synthesis proceeds along the minus-strand of the vDNA, thereby copying the U5 region and PBS on to the plus-strand of vDNA creating the LTR, while RNase H degrades the PPT and the

tRNA primer. The degradation of the tRNA primer enables the second jump or plus-strand transfer. Both the plus- and minus-strands are extended until the entire vDNA is double-stranded (reviewed in, Hughes, 2015).

Uncoating of the capsid core is thought to be a highly regulated process. Previously, it was suggested that the process of uncoating is linked to reverse transcription (Arhel et al., 2007, Hulme et al., 2011), triggered by the synthesis of first strand transfer during reverse transcription (Cosnefroy et al., 2016). However, more recent reports suggested that uncoating begins after docking to a nuclear pore (Francis and Melikyan, 2018) or within the nucleus, prior to the completion of reverse transcription (Francis et al., 2020), or immediately before the integration (Li et al., 2021).

At the completion of reverse transcription, retroviral DNA becomes part of a large nucleoprotein complex, termed the pre-integration complex (PIC) (Bowerman et al., 1989). Within the PIC, vDNA is associated with a range of viral proteins, most notably IN, and likely other viral components (Miller et al., 1997) as well as host proteins, including lens epithelium-derived growth factor p75 splice variant (LEDGF/p75) (Llano et al., 2006, Vandegraaff et al., 2006). Retroviral IN catalyses the integration reaction, where a pair of vDNA ends are covalently and irreversibly attached to the host chromosome, forming a stable provirus.

The integrated proviral DNA is transcribed to generate a pre-mRNA that undergoes alternative splicing, thereby creating multiple spliced mRNAs (Purcell and Martin, 1993). These can be broadly classified into completely or multiply spliced, singly or partially spliced and unspliced mRNA species. The unspliced RNA transcripts are about 9-kb long and form the template for translation of Gag and Gag-Pol. Crucially, two copies of the unspliced RNA transcript are packaged into assembling viral particles to serve as the genomic RNA in the next infection cycle (Cullen, 1998, Kuzembayeva et al., 2014). The incorporation of genomic RNA is an important step in the viral particle assembly and depends on the specific interaction between the packaging signal within viral RNA and NC (Wang and Aldovini, 2002). Immature HIV-1 particles are assembled by Gag-Gag interactions into a curved lattice at the cell plasma membrane. HIV-1 release depends on action of the endosomal sorting complex required for transport (ESCRT) machinery, facilitated by Gag p6 interaction with an ESCRT protein (Garrus et al., 2001, Martin-Serrano et al., 2001). Following

viral release, activated viral protease cleaves Gag and Gag-Pol polyproteins at specific positions, leading to the formation of a mature virus particle (Pettit et al., 1994, Kräusslich et al., 1995, Wiegers et al., 1998, Mattei et al., 2016b). HIV-1 maturation is accompanied by transformation of the CA lattice into conically shaped capsids, encasing viral RNA in complex with NC and IN (Briggs et al., 2004, Briggs et al., 2006, Kessl et al., 2016).

1.5 Retroviral integrase (IN) structure and functions

1.5.1 **IN domain organisation**

All retroviral IN proteins harbour three conserved domains: the N-terminal domain (NTD), the catalytic core domain (CCD) and the C-terminal domain (CTD), which are joined by highly variable linkers (Figure 1-4). In addition, spumaviral, epsilon and gammaretroviral INs possess an N-terminal extension domain (NED) that precedes the NTD (Hare et al., 2010, Lesbats et al., 2016, Guan et al., 2017). The CCD is the most conserved domain (Khan et al., 1991), while the CTD is least conserved by amino acid sequence (Lodi et al., 1995).

Solution structure of HIV-1 IN NTD was initially resolved in its dimeric form by nuclear magnetic resonance (NMR) spectroscopy (Cai et al., 1997). The NTD harbours the invariant HHCC motif, residues of which participate in coordination of a Zn^{+2} ion (Burke et al., 1992); the domain forms a highly stable helical bundle stabilised by Zn^{+2} binding. The metal ion is coordinated tetrahedrally by His-12, His-16, Cys-40 and Cys-43 (Cai et al., 1997). This domain is essential for tetramerisation of HIV-1 IN (Zheng et al., 1996, Cai et al., 1997, Hare et al., 2009a). Disruption of the NTD structure via H12N or H16N mutations substitutions grossly affected HIV-1 IN catalytic activities *in vitro* (Engelman and Craigie, 1992). The NTD is composed of four α-helices (Figure 1-4C). Within a dimer observed in solution, the NTDs are arranged in parallel to each other. N-terminal ends of helices α1, α3 and α4 make intermolecular contacts within the dimer. The structure of the NTD is similar to the helix–turn–helix DNA-binding domains, such as the Trp repressor (Otwinowski et al., 1988). One of the helices that contacts DNA in the Trp repressor forms part of the dimer interface in the case of the NTD of HIV-1 IN (Cai et al., 1997).

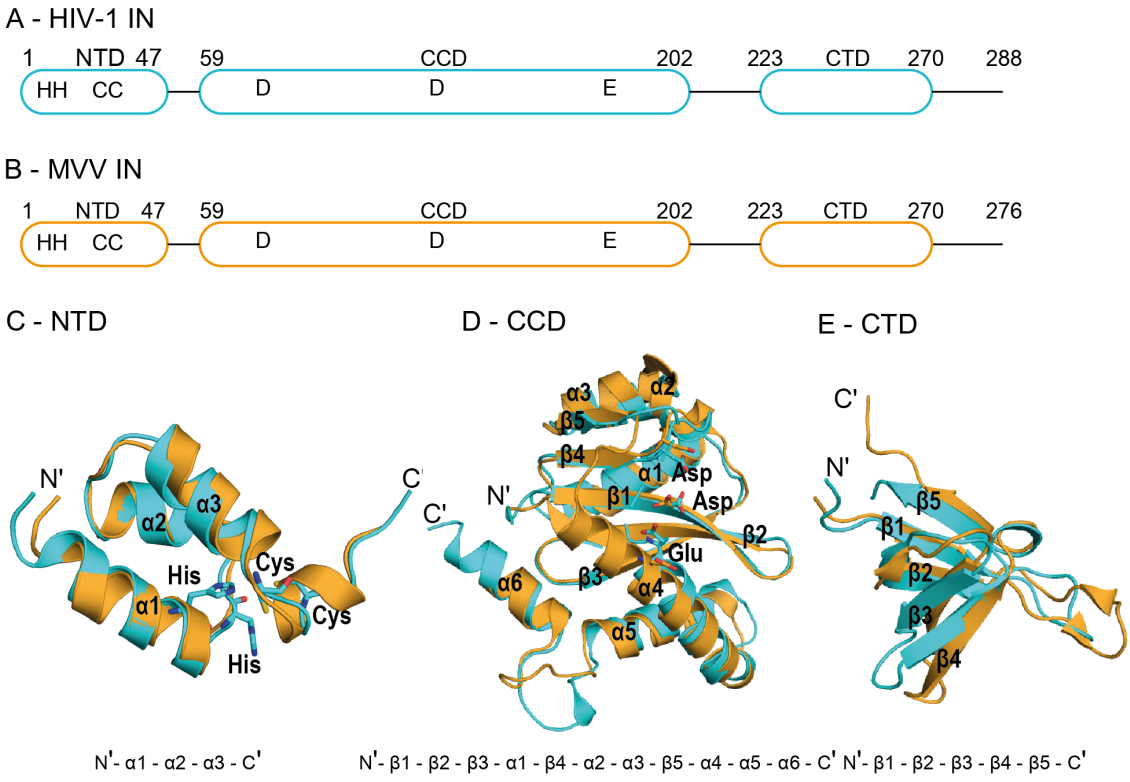

Figure 1-4 Domain organisation of HIV-1 and MVV INs.

(A) HIV-1 and (B) MVV INs comprise the NTD, hosting the invariant HHCC motif, CCD, harbouring the enzyme active site motif D,D-35-E and the CTD. (B, C &D) Structures of the respective HIV-1 and MVV IN domains superimposed, HIV-1 (cyan, PDB IDs: 1WJC, 1BIS, 1QMC) and MVV IN (gold, PDB IDs: 3HPG, 5T3A and 5LLJ), respectively. The order of α-helices and β-strands (N′ to C′) is indicated below the domain structures. The domain features, including α-helices and β-sheets are indicated. The invariable HHCC motif in the NTD and DDE motif in the CCD are indicated by sticks.

The CCD hosts the invariant D,D-35-E motif (Khan et al., 1991, Kulkosky et al., 1992) that coordinates a pair of Mg^{2+} ions to form the enzyme active site for performing the reactions, 3′-processing and strand transfer. Although Mg^{2+} is thought to catalyse the reaction *in vivo*, due to its availability, this metal ion can be substituted with Mn^{2+} *in vitro*, while retraining the IN activity (Engelman and Craigie, 1995, Hazuda et al., 1997). Amino acid substitutions within the HIV-1 IN D,D-35-E motif, including D116N and E152Q, abrogated all catalytic activities *in vitro* (Engelman and Craigie, 1992). In the context of viral infection, alanine substitutions of HIV-1 IN Asp-64, Asp-116 and Glu-152 led to the formation of circular forms of vDNA and were defective for integration (Wiskerchen and Muesing, 1995, Leavitt et al., 1996). Recombinant HIV-1 IN and most of its deletion constructs spanning the CCD are poorly soluble. A single Phe -to-Lys substitution at HIV-1 IN position 185 significantly improved the protein's solubility allowing the initial structural characterisation of the CCD by X-ray crystallography (Dyda et al., 1994). The structure revealed a dimer with subunits oriented in parallel to one another, with each subunit composed of five β-strand sheets and six α-helices (Figure 1-4D) (Dyda et al., 1994). Structural similarity of the CCD was noted with RNase H (Davies et al., 1991), the Holliday junction resolvase (Ariyoshi et al., 1994) and the core domain of Mu phage transposase protein (Dyda et al., 1994).

The HIV-1 IN CTD structure was initially determined by NMR spectroscopy (Lodi et al., 1995, Eijkelenboom et al., 1995). The monomer is composed of five antiparallel β-strands, forming a five-stranded β-barrel, characterized as an SH3-like fold (Figure 1-4E). In solution, the CTD forms a stable dimer, which is formed by a stacked β-interface involving β2, β3, and β4 strands with the two three-stranded anti parallel β-sheets, one from each subunit, arranged in an antiparallel orientation. The pronounced DNA binding activity of the IN CTD is driven by the properties of specific amino acid residues. In particular, IN CTD residues Leu-234 and Arg-262 were shown to bind DNA through their hydrophobic interactions and electrostatic interactions, respectively. HIV-1 IN CTD substitutions R231A, L234A, L241A, L242A, V260A and R262G had reduced 3′-processing and strand transfer activities *in vitro* (Lutzke and Plasterk, 1998).

The linkers connecting the three canonical IN domains display very little sequence conservation and are thought to be highly flexible. The only exception appears to be

the CCD-CTD linkers in lentiviral INs, which assume a relatively rigid α-helical conformation (Chen et al., 2000, Ballandras-Colas et al., 2017). HIV-1 IN and MVV IN are composed of 288 and 276 amino acid residues, respectively (Figure 1-4A and B). The nomenclature of the amino acid residues in HIV-1 and MVV INs is based on their positions in full-length mature forms (*i.e*. after cleavage of the Gag-Pol precursor).

1.5.2 **The process of retroviral integration**

The vDNA molecule, synthesised during reverse transcription, becomes a part of the PIC together with a range of cellular and viral proteins (Bowerman et al., 1989, Miller et al., 1997). Most notably, a multimer of retroviral IN subunits assemble on vDNA ends, forming a highly stable complex, termed the intasome. The intasome catalyses insertion of the vDNA into chromosomal DNA. To facilitate the integration process, retroviral IN performs two essential catalytic steps: 3′-processing and strand transfer. During 3′-processing, it hydrolyses a phosphodiester bond at both the vDNA ends, to expose 3′-hydroxyl groups (3′-OH) attached to the invariant cytosine-adenine (dCdA) dinucleotides, forming a cleaved synaptic complex (Figure 1-5) (Bushman and Craigie, 1991, Engelman et al., 1991). The intasome then binds to the host chromosomal DNA, forming target capture complex (TCC) and processed 3′ ends of the vDNA are covalently linked to the target DNA during strand transfer (Figure 1-5). Next, the 5′ vDNA overhangs generated during 3′-processing are removed and single stranded gaps are repaired by the host DNA repair machinery (Yoder and Bushman, 2000), resulting in a fully integrated provirus (Figure 1-5).

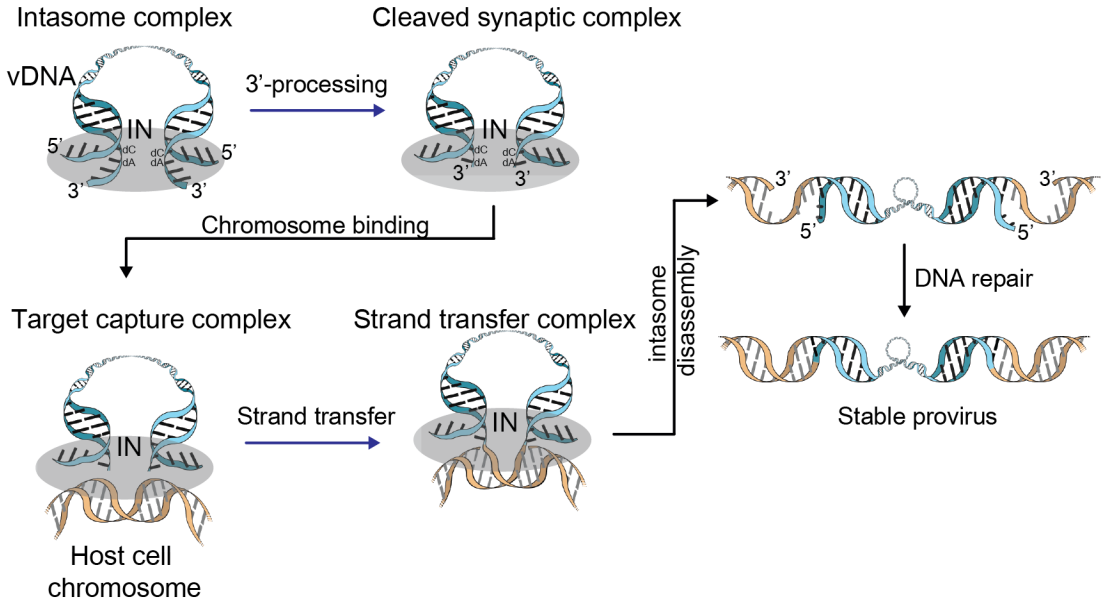

Figure 1-5 Illustration of the steps involved in the retroviral integration process.

A multimer of IN subunits (grey oval) assemble at the vDNA ends, forming an intasome. During 3′-processing, nucleotides downstream of the invariant CA dinucleotide are cleaved, forming a cleaved intasome complex. It then binds to the target DNA, forming the target capture complex (TCC). The processed vDNA ends are covalently attached to the target DNA, forming strand transfer complex (STC). The 5′ overhangs are removed, and single stranded gaps are repaired by the host DNA repair machinery, thereby establishing a provirus (Figure adapted from Lesbats et al., 2016, Hare et al., 2012).

1.5.3 **The mechanism of retroviral integration**

The first intasome to be structurally characterized was assembled using recombinant WT PFV IN and 19-bp double-stranded oligonucleotides mimicking 3′-processed U5 vDNA ends. Its high resolution three-dimensional structure was determined by X-ray crystallography (Figure 1-6) (Hare et al., 2010). This advancement enabled detailed analysis of the integration reaction (Maertens et al., 2010, Hare et al., 2012).

The PFV intasome is composed of a homotetramer of IN (dimer-of-dimers) in association with a pair of vDNA ends (Figure 1-6). Within the intasomal tetramer, the inner IN subunits form all the contacts with vDNA and provide the active sites for catalysis. The outer IN subunits interact with the CCDs of the inner chains and were later shown to play a role in integration into chromatinized target DNA (Maskell et al., 2015), elaborated in section 1.7. The NTDs interact with the inner CCDs in *trans*. The CCD-CTD linker is in an extended conformation and runs in parallel with the NTD-CCD linker from the same subunit. The intasome is stabilised by insertion of a pair of CTDs between the two catalytic CCDs (Hare et al., 2010), as shown in (Figure 1-6). Because they stabilise the intasomal synaptic interface, they are referred to as "synaptic CTDs" to discriminate them from the CTDs that play other functions within the more complex alpha-, betaretroviral and lentiviral intasomes (as explained in section 1.5.4).

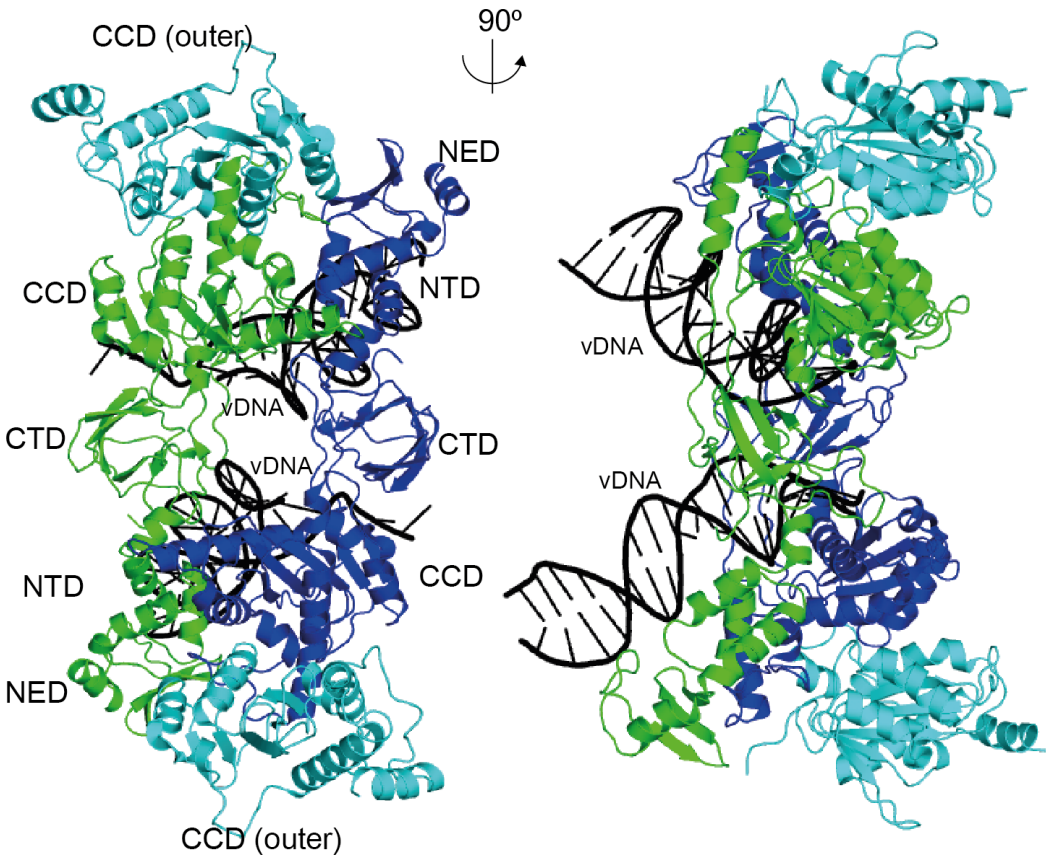

Figure 1-6 PFV intasome structure.

Views along (left) and perpendicular to (right) the crystallographic two-fold axis of the PFV intasome (PDB ID 3OY9). The inner IN subunits (coloured in dark blue and green), with the NTDs interacting with CCDs in *trans*, while the CTDs are placed in the synaptic position. The vDNA (black) passes through the centre of the complex. The outer CCDs are coloured in cyan (Hare et al., 2010, Hare et al., 2009a).

All PFV IN domains, excluding the NED, and the inter-domain linkers of the inner subunits interact in a sequence-specific manner with the vDNA bases, present on both the reactive vDNA strand (*i.e*. the strand that undergoes 3′-processing and strand transfer) as well as the non-reactive vDNA strand. The terminal base pairs of vDNA are separated by the CCD α4 helix of the IN. PFV IN Gly-218 and Arg-222 form hydrogen bonds with guanine 4 (dG-4) and with thymine 5 (dT-5) and dC-6 bases of the non-reactive vDNA strand, respectively (Figure 1-7A), Asn-106 forms a hydrogen bond with dT-8 of the non-reactive strand (Figure 1-7B). The side chain of Arg-313 interacts with dA-12 of the reactive vDNA strand and forms a hydrogen bond with dC-11 (Figure 1-7B). The vDNA 3′ ends, with the invariant dCdA dinucleotide (dC-16 and dA-17), are placed within reach of the active site, composed of Asp-128, Asp-185 and Glu-221, comprising the D,D-35-E motif in PFV IN (Hare et al., 2010). Subsequently, the two catalytic steps involved in the integration reaction were visualised at an atomic level by crystallography of the PFV intasome complexes along with or without the target DNA (Maertens et al., 2010, Hare et al., 2012).

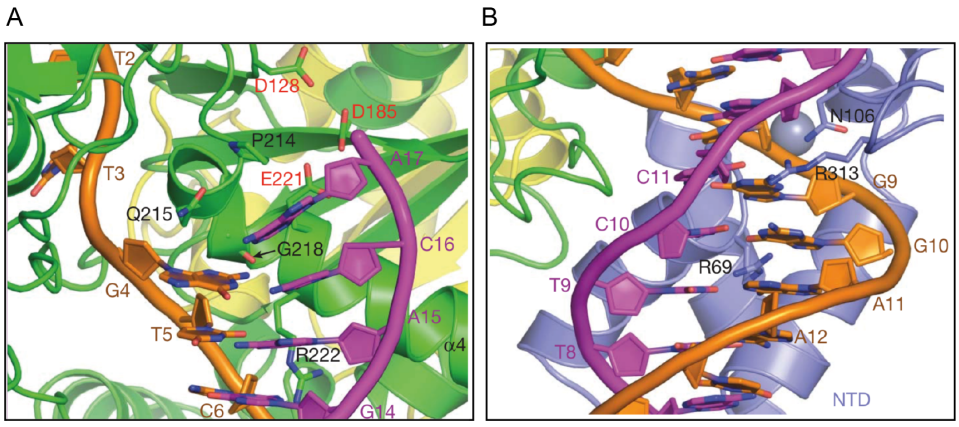

Figure 1-7 Interactions of IN-vDNA within PFV intasome.

Interactions of amino acid residues and vDNA in PFV intasome (PDB ID 3OY9) discussed in text. (A) Gly-218 and Arg-222 interactions with DNA bases of non-reactive strand (orange); (B) interactions of Asn-106 and Arg-313 with DNA bases from both reactive strand (magenta) and non-reactive strands of vDNA. Figure from (Hare et al., 2010), with appropriate permissions from Springer Nature.

PFV intasome was assembled with blunt-ended unprocessed vDNA, which was crystallised, and the catalytic step of 3′-processing was visualised by supplementing the crystals with Mg^{2+}or Mn^{2+} salts (Hare et al., 2012). Two metal ions (traditionally referred to as metal ions A and B), are coordinated between the acidic residues of the D,D-35-E active site triad. During 3′-processing, both the metal ions present in octahedral coordination with the enzyme active sites, interact with the targeted phosphodiester bond on vDNA (present between dA-17 and dA-18, Figure 1-8). This interaction enables the placement of the scissile phosphate group for an in-line S_N2 nucleophilic substitution by a water molecule (labelled as W_{Nuc} in Figure 1-8) that is coordinated by metal ion A. Following hydrolysis, the scissile dinucleotide dissociates, leaving 3′-OH attached to dA-17 (Hare et al., 2012). This sets the stage for the strand transfer of the processed vDNA on to the target DNA.

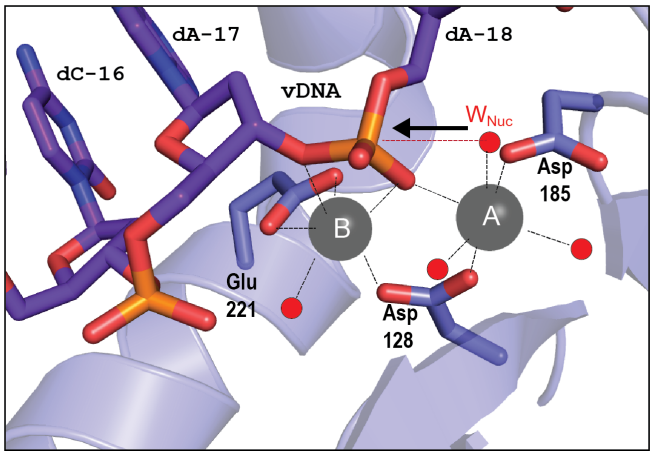

Figure 1-8 3'-processing performed by PFV intasome visualised at the atomic level.

Crystal structure of the PFV intasome (PDB ID 4E71) capturing the positioning of the active sites (Asp-128, Asp-185 and Glu-221, indicated in sticks) and divalent metal ions (A and B, indicated in black spheres) with the scissile nucleotide of the reactive vDNA (purple) prior to 3′-processing. The metal ions are octahedrally coordinated (black dashed lines) with the residues of the active site and four water molecules (red circles). Nucleotides downstream of dA-17 are removed in an S_N2 nucleophilic substitution, where a water molecule (W_{Nuc}) acts as a nucleophile, coordinated by metal ion A and attacks the phosphodiester bond between dA-17 and dA-18 (Hare et al., 2012, Hare et al., 2009a).

Co-crystallisation of the PFV intasome with a short double-stranded oligonucleotide mimicking target DNA and incubating the crystals in the presence of Mn^{2+}or Mg^{2+} ions allowed strand transfer to occur within crystals. Brief soaks followed by freezing the crystals in liquid nitrogen made it possible to visualise the TCC committed for reaction, with both metal ions A and B in the intasomal active site (Maertens et al., 2010, Hare et al., 2012) (Figure 1-9).

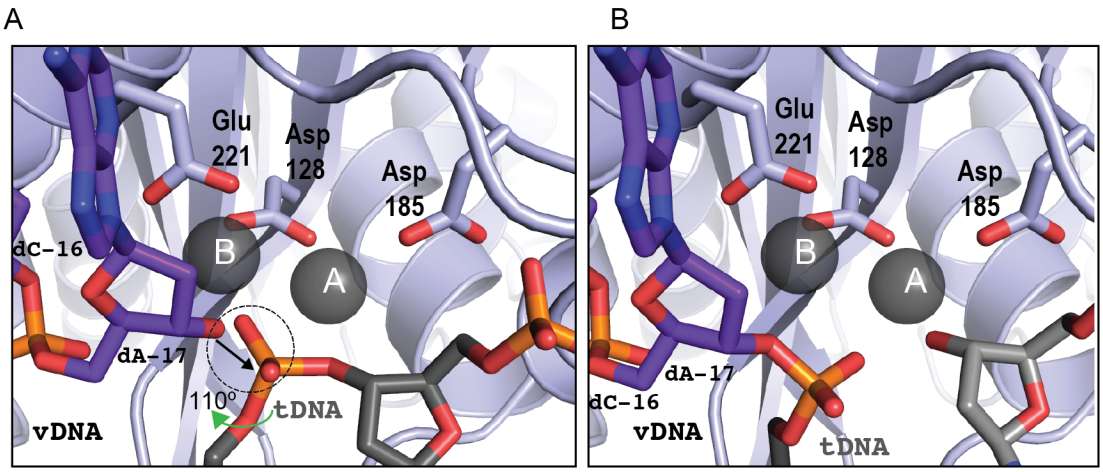

Figure 1-9 Strand transfer reaction performed by the PFV intasome captured *in crystallo*.

Crystal structure of the PFV intasome (A) target capture complex (TCC, PDB ID 4E7K) and (B) strand transfer complex (STC, PDB ID 3OS0) capturing the positioning of the active sites (Asp-128, Asp-185 and Glu-221, indicated in sticks) and divalent metal ions (A and B, indicated in black spheres) with the scissile phosphodiester bond on the target DNA (graphite) before and after strand transfer. The 3′-OH group of the processed vDNA (purple) is placed in-line with the phosphate group of the target DNA, coordinated by metal ion B, for S_N2 nucleophilic substitution, which enables the strand transfer reaction to occur on to the target DNA (Maertens et al., 2010, Hare et al., 2009a).

The target DNA is sharply bent when in complex with the intasome and is bound between the two halves of the intasome. Deformation of target DNA facilitates positioning of its scissile phosphodiester groups for strand transfer. Metal ion B coordinates the positioning of the 3′-OH group of 3′-dA of vDNA for S_N2 nucleophilic substitution reaction at the phosphorus atom in target DNA (Figure 1-9A). Following strand transfer (visualised after a longer incubation of the crystal in Mn^{2+}), a 110º rotation of the sugar moiety of target DNA at the point of strand transfer was seen (Figure 1-9A). This conformation change allows the shifting of the newly formed phosphodiester bond between vDNA and target DNA away from the active site (Figure 1-9B) and is thought to make the S_N2 substitution irreversible (Maertens et al., 2010, Hare et al., 2012).

1.5.4 **Other retroviral intasomes**

A number of retroviral intasome structures have been reported in the past 10 years. The intasome complexes assembled using deltaretroviral (HTLV-I and STLV) INs were recently determined by cryo-electron microscopy (cryo-EM) in complex with their host factor, the regulatory subunit B56γ of human serine/threonine-protein phosphatase 2A (PP2A) (Barski et al., 2020, Bhatt et al., 2020). Similar to the PFV intasome assembly (Figure 1-10A), the deltaretroviral intasomes harbour a tetramer of IN. The NTDs of the catalytic IN subunits interact with the CCDs *in trans*. Two B56γ subunits present in the complex that interact with two IN subunits each on either side in the intasome complex (Figure 1-10B). A unique role for the IN CCD-CTD linkers was reported, where they had two binding sites for the B56γ subunits (Barski et al., 2020).

The intasome structures of alpharetroviral (ASLV) and betaretroviral (MMTV) intasomes were also determined (Yin et al., 2016, Pandey et al., 2021, Ballandras-Colas et al., 2016). The ASLV intasome was assembled with catalytically active IN and a branched DNA mimicking the product of strand transfer, which was then crystallised (Yin et al., 2016). In contrast, MMTV intasome was assembled with recombinant MMTV IN and vDNA, which was then analysed by cryo-EM (Ballandras-Colas et al., 2016). Both the ASLV and the MMTV intasomes featured octameric IN assemblies, each composed of four IN dimers. Within the intasome, the two core

dimers provided one active site each, while the two flanking IN dimers each contributed a synaptic CTD (Figure 1-10C).

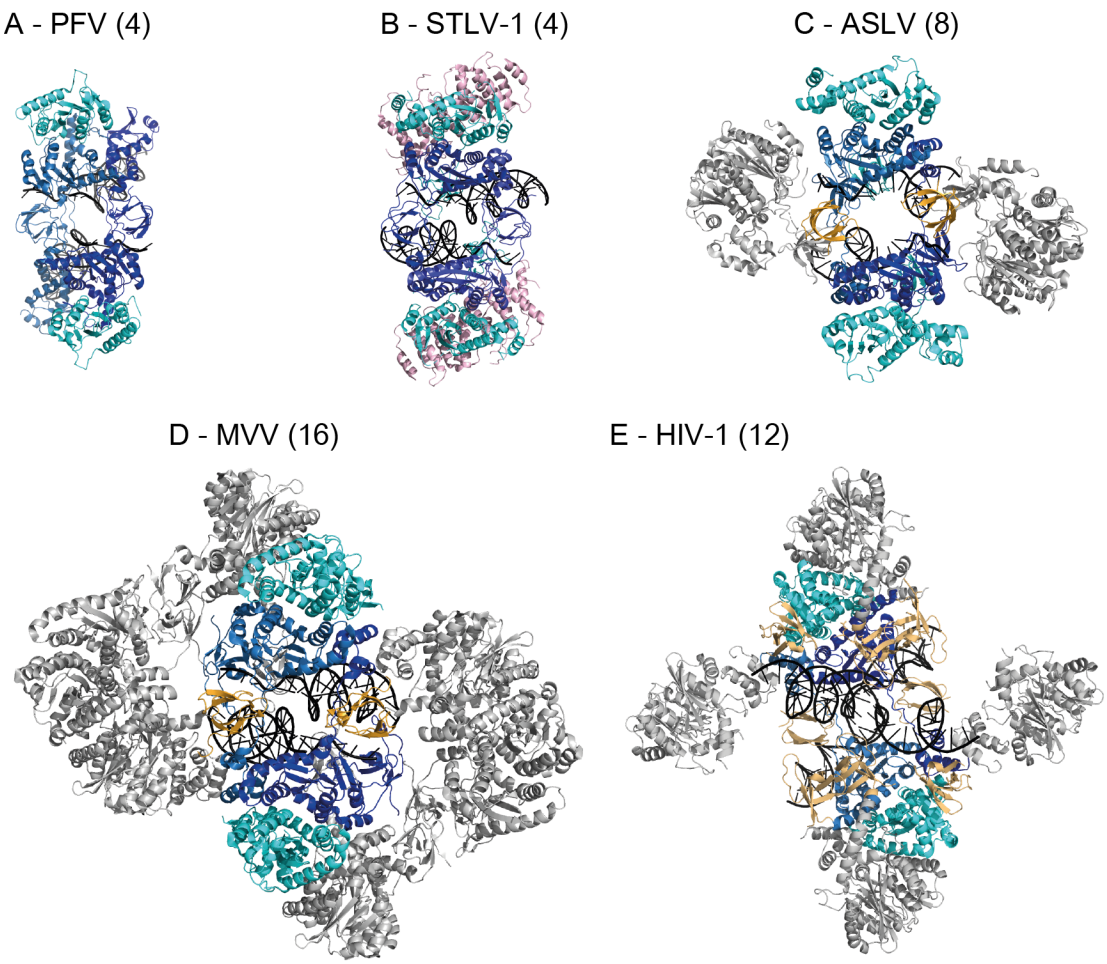

Figure 1-10 Structures of the known retroviral intasomes.

The intasome structures of retroviruses have their CIC's highlighted in blues, with CTDs contributed by flanking IN subunits, in case of ASLV, MVV, HIV-1 and SIV intasomes, in gold. The IN subunits not part of the CIC are in grey. (A) PFV intasome, a tetrameric assembly (PDB ID 3OY9); (B) STLV-1 intasome, with B56γ subunits in pink, a tetrameric assembly (PDB ID 6Z2Y) (C) ASLV intasome, an octameric assembly (PDB ID 5EJK); (C) MVV intasome, a hexadecameric assembly (PDB ID 5M0Q) and (D) HIV-1 intasome, partial dodecameric assembly (PDB ID 5U1C).

The lentiviral MVV intasome was assembled with recombinant WT IN, vDNA oligonucleotide and host factor of lentiviral integration, LEDGF/p75. Its structure was determined by cryo-EM, revealing sixteen MVV IN subunits in the resulting cryo-EM map (Figure 1-10D, 1-11A). The hexadecameric MVV intasome is the largest assembly of its kind observed to-date. Four IN tetramers assembled into the intasome, of which two core tetramers contributed one active site each and two flanking tetramers contributed a synaptic CTD each (Ballandras-Colas et al., 2017). The MVV intasome is roughly four times larger than the equivalent PFV assembly. Within the MVV intasome structure, both the vDNA ends, with 20 bp of vDNA, are clearly defined. The non-reactive vDNA strand is separated from the reactive strand and the processed vDNA end is situated at the active site of the core IN (subunits A and I) (Figure 1-11A). Similar to the PFV intasome, the NTDs from the catalytic subunits are exchanged between each other, with the NTD-CCD linkers passing along the synaptic interface (Ballandras-Colas et al., 2017). The MVV STC structure determined by cryo-EM (Ballandras-Colas et al., 2017, Ballandras-Colas et al., Manuscript in preparation) agrees with the prior PFV (Maertens et al., 2010) and ASLV STC structures (Yin et al., 2016). The target DNA is severely bent and fits into the crevice at synaptic interface, which is caused by the insertion of synaptic CTDs β1-β2 loops into expanded major grooves (Figure 1-11C).

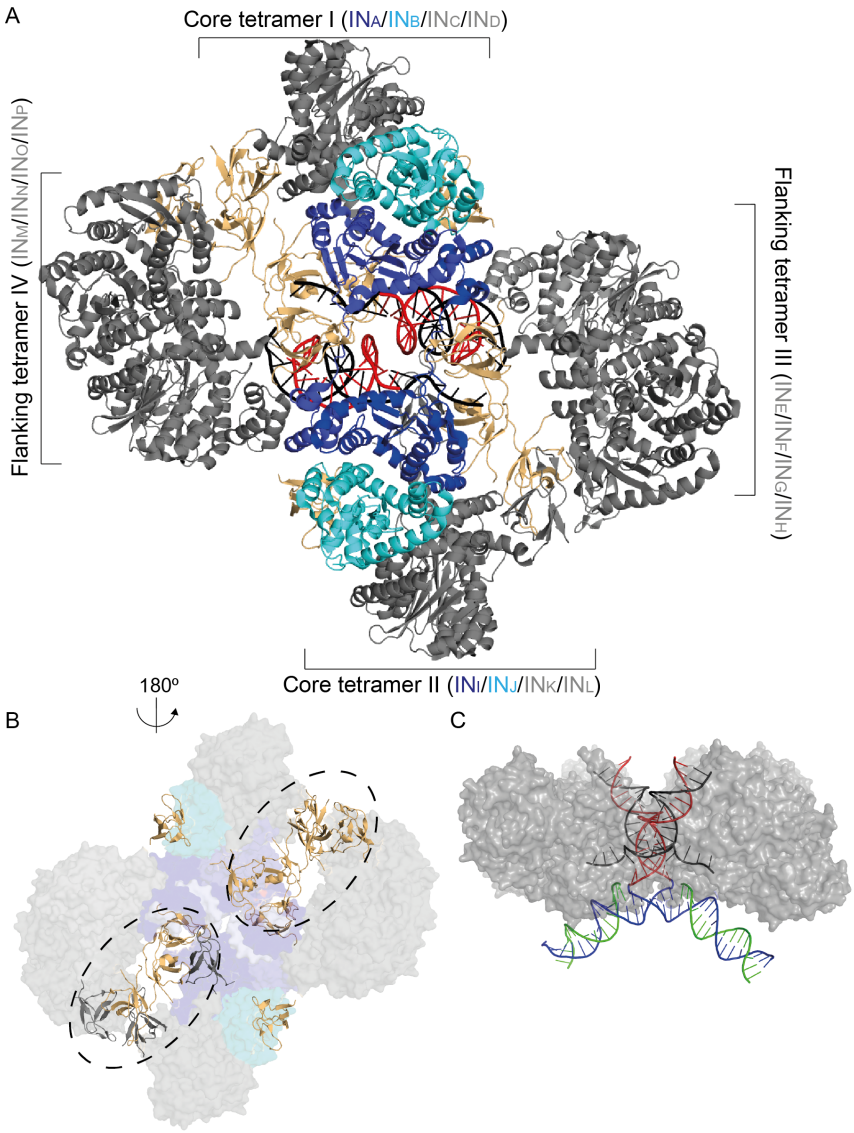

Figure 1-11 MVV intasome structure.

(A) The MVV intasome structure (PDB ID 5M0Q) is composed of a tetramer of tetramers, with two core tetramers I and II and two flanking tetramers III and IV. The IN subunits part of the CIC are in shades of blue with the additional IN subunits in grey, the CTDs in gold and reactive vDNA strands in red and non-reactive vDNA strands in black. (B) The intasome rotated by 180º, only the CTDs shown as cartoon and the CTD bridge indicated by black circles. (C) The cryo-EM STC structure (5M0R), the protein in grey and scissile target DNA strands in blue and unused target DNA strands in green (Ballandras-Colas et al., 2017).

Assembly of the HIV-1 intasome required drastic modifications to the IN protein due to the unfavourable biochemical properties of WT HIV-1 IN, such as poor activity and low solubility of recombinant protein. HIV-1 IN N-terminal fusion with *Sulfolobus solfataricus* chromosomal protein, Sso7d was described previously to enhance the biochemical properties of the protein *in vitro* (Li et al., 2014). Sso7d binds to DNA in a sequence-independent manner and its fusion to DNA polymerases enhanced their activities *in vitro* (Wang et al., 2004). The Sso7d-IN chimera was used to assemble the intasome, along with branched DNA, mimicking full-site integration product, and the integrase-binding domain (IBD) of LEDGF/p75, the host factor (see section 1.5.6.3). This work resulted in a collection of HIV-1 intasome structures determined by cryo-EM (Figure 1-10E) (Passos et al., 2017). The largest HIV-1 intasomal assembly contained twelve IN subunits. This heterogeneity is likely a consequence of the presence of Sso7d domain, which could hinder the assembly of higher order multimers observed with the MVV intasome (Ballandras-Colas et al., 2017). While missing two IN dimers, the dodecameric HIV-1 intasome can be fully superposed on the hexadecameric MVV structure (Ballandras-Colas et al., 2017, Engelman and Cherepanov, 2017, Passos et al., 2017).

Recently, the structures of red-capped mangabey SIV (SIV_{rcm}) intasome were determined by cryo-EM to a near-atomic resolution (Cook et al., 2020). The intasome was assembled with recombinant full-length WT SIV_{rcm} IN, LEDGF/p75 and substrate vDNA mimicking the viral U5 ends. Despite heterogeneity of the SIV_{rcm} intasome samples, negative-stain EM analysis revealed the presence of hexadecameric species reminiscent of the MVV intasomes, along with various partial assemblies and artefactual intasome polymer stacks. As observed with MVV, SIV_{rcm} intasome harbours two core IN tetramers, which engage vDNA ends and a pair of CTDs donated by flanking subunits (Cook et al., 2020).

Heterogeneity of the primate lentiviral (HIV-1 and SIV_{rcm}) intasomes prevent isolation and study of single monodisperse species (Passos et al., 2017, Cook et al., 2020, Passos et al., 2020). The MVV intasome structure was determined from an assembled complex with the WT form of the recombinant IN protein, and the intasome particles existed in near-monodispersed population with a predominance of hexadecameric units of INs, allowing for the deciphering of complete intasome structural information (Ballandras-Colas et al., 2017). The distinctive features

observed in the MVV intasome, are also present in the available HIV-1 and SIV_{rcm} structures. These features include, intra-tetramer CTD-CTD interactions (Figure 1-11B) and, as determined by the lentiviral IN's CCD-CTD crystal structures, the α-helical linker (Chen et al., 2000, Ballandras-Colas et al., 2017). Thus, MVV intasome appears to be an ideal model for dissecting lentiviral integration *in vitro*. Hence, the main aim of this study was to test the biological significance of the expanded intasome MVV assembly.

1.5.5 The conserved intasome core (CIC)

Despite their striking and unexpected diversity, all intasomes share the conserved intasome core (CIC), which harbours two CCD dimers joined by exchange of the NTDs and a pair of synaptic CTDs. Each CCD dimer within the CIC contributes one active site (Figure 1-12). The tetrameric PFV intasome is the minimalist assembly, representing little more than the CIC (the PFV IN NEDs, are absent in INs from most retroviral species, and are therefore not considered to be parts of the CIC). The intasome structures from other retroviruses contain the CICs decorated by variable amount of additional IN subunits.

In solution, PFV INs are monomers, ASLV and MMTV are dimers, while lentiviral INs are at least tetramers. Thus, the intasomes seem to assemble by tetramerisation of the respective multimeric subunits of INs in solution. In PFV intasome, the synaptic CTDs are contributed by the core IN subunits (Figure 1-12A). In ASLV and MMTV, flanking dimers contribute a CTD each to the core dimers (Figure 1-12B) and similarly, in MVV, flanking tetramers contribute a CTD each to the core tetramers (Figure 1-12C). In theory, as seen in PFV intasome, the CIC alone should be sufficient to support integration.

One of the explanations for the relatively large sizes of intasome assemblies in orthoretroviral INs is the lengths of the CCD-CTD linkers in the respective IN proteins. In PFV IN, this linker spans 60 amino acid residues, without a defined secondary structure. This long polypeptide chain enables the CTD from the catalytic subunit to attain an extended configuration and reach the synaptic position (Figure 1-12A). By sharp contrast, in alpha- and betaretroviruses, the CCD-CTD linker is less than 10 amino acid residues long, thereby preventing the CTD from the catalytic subunit to

be part of the CIC. This requires additional dimers to contribute their CTDs for the synaptic positions (Figure 1-12B). Although the lentiviral CCD-CTD linker is of an intermediate length of 20 amino acid residues, similar to that of the deltaretroviral INs (Barski et al., 2020), it uniquely exists in a compact α-helical configuration (Chen et al., 2000, Ballandras-Colas et al., 2017). This necessitates the presence of the two flanking IN tetramers to contribute the synaptic CTDs complete the CIC (Figure 1-12C). These considerations provide a structural explanation of the constraints that necessitate a larger intasome in lentiviruses than other retroviral species. However, it does not provide any insight into the functional consequences of the intasome expansion in lentiviruses. This leads to the question of whether a functional lentiviral intasome can be assembled with reduced number of IN subunits and what the potential advantages are of this larger assembly for lentiviral integration.

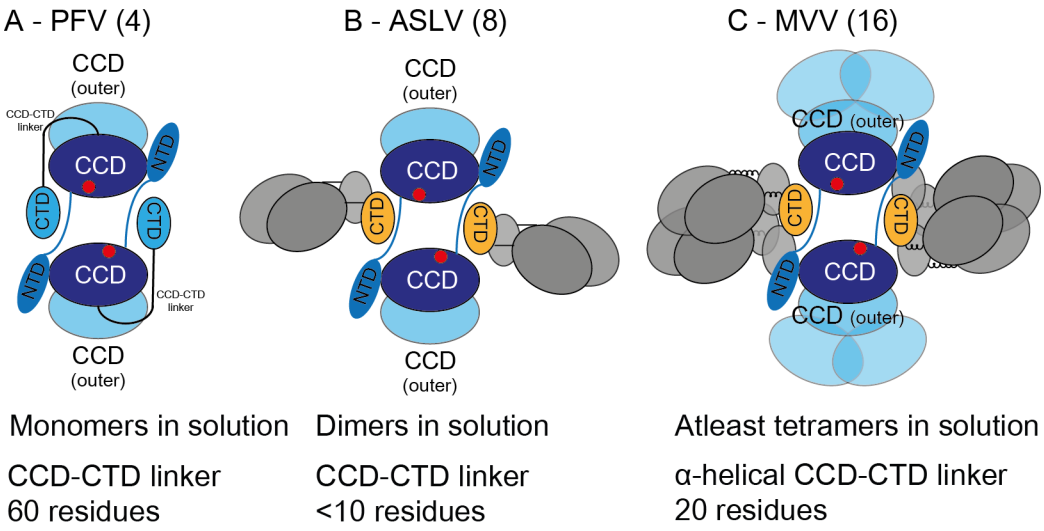

Figure 1-12 Cartoon illustrations of the PFV, ASLV and MVV intasomes.

The CIC's highlighted in shades of blue, with CTDs contributed by flanking IN subunits, in case of ASLV and MVV in gold. The IN subunits not part of the CIC are in grey. (A) PFV intasome, a tetrameric assembly; (B) ASLV intasome, an octameric assembly; (C) MVV intasome, a hexadecameric assembly. Their multimeric states and CCD-CTD linker lengths are indicated.

1.5.6 Intasome within the PIC

Early on in the retroviral field, prior to the determination of the intasome structures, biochemical analyses of retroviral PIC purified from infected cells were carried out by Mu phage transposase-mediated DNA foot-printing. The transposase (Mu A protein) was used to insert Mu-DNA into MLV vDNA (Wei et al., 1997) and HIV-1 vDNA (Chen et al., 1999) within PICs extracted from infected cells. These studies reported an extensive protection of both vDNA ends from the attack by the transposase. This ranged from ~200 up to ~500 nucleotides from the vDNA ends into the viral genome and not present in the mid-region of the viral genome. The interpretation of these results is that an extensive protein complex assembles on vDNA ends within the retroviral PIC structure.

The variable IN multimers within retroviral intasomes cover <20 bp at each vDNA end, suggesting that the native complex that assembles on vDNA ends *in vivo* is much larger. Although the exact composition of the PIC is still unknown, several studies were carried out to identify the proteins involved. In addition to IN, the presence of a number of viral proteins was detected in the PIC, including CA, RT, MA, and Vpr (Bukrinsky et al., 1993, Heinzinger et al., 1994, Gallay et al., 1995, Miller et al., 1997, Yamashita and Emerman, 2004). The PICs were also reported to contain numerous cellular proteins, including, but not limited to, high-mobility group protein 1A (HMG1A) (Farnet and Bushman, 1997), barrier-to-autointegration factor (BAF) (Lin and Engelman, 2003), LEDGF/p75 (Llano et al., 2004), histone-binding partner RBBP4 (Raghavendra et al., 2010) and annexin A6 (ANXA6) (Schweitzer et al., 2013).

1.5.6.1 HMG1A

The high mobility group (HMG) proteins were characterised as the most abundant non-histone DNA-binding proteins in mammalian cells, with HMGA proteins in particular containing AT-hooks (Bustin and Reeves, 1996). These proteins were implicated in modifying the DNA structure to aid in protein-protein interactions within the transcription preinitiation complex, particularly formed in an A/T-rich promoter sequences of genes (Fashena et al., 1992). The presence of HMG1A was detected in HIV-1 PICs purified from infected cells, and these proteins were removed on high-

salt treatment (600 mM KCl). Integration assays were carried out with the salt-treated PICs in the presence of HMG1 proteins, which restored the integration activity to a large extent. These proteins were hence characterised as accessory factors for the integration function of HIV-1 PICs (Farnet and Bushman, 1997).

1.5.6.2 Barrier-to-autointegration factor (BAF)

Nucleoprotein complexes containing vDNA purified from MLV-infected cells had the ability to prevent autointegration of vDNA, which was disrupted upon salt treatment and restored on treating with host cell extracts (Lee and Craigie, 1994). The cellular factor responsible for this protection was identified as the 89-amino acid residue protein BAF (Lee and Craigie, 1998). It was later shown to restore the integration activity of salt-treated HIV-1 PICs and was also found to be a host factor for HIV-1 integration (Chen and Engelman, 1998). BAF is a DNA-binding protein, which induces condensation of DNA (Skoko et al., 2009). Although the precise mechanism of BAF interaction with the vDNA and its necessity for lentiviral infection have not been determined, it was suggested to play a role in compaction of vDNA within the PIC, which potentially aids in preventing autointegration of the vDNA (Skoko et al., 2009).

1.5.6.3 Lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75)

LEDGF/p75 is by far the most validated host factor involved in HIV-1 (and generally lentiviral) integration (Bedwell and Engelman, 2021). Immunoprecipitation of FLAG-tagged HIV-1 IN from extracts of HEK-293T cells ectopically expressing the viral protein revealed a cellular protein bound to it (Cherepanov et al., 2003). This was identified as LEDGF/p75 by mass spectrometry. Despite its true molecular mass of ~60 kDa, LEDGF/p75 migrates at 75 kDa in SDS-PAGE gels. Also, it was observed that LEDGF/p75 co-localises with HIV-1 IN in the nucleus of HEK-293T cells. Moreover, in the presence of LEDGF/p75, the strand transfer activity of HIV-1 IN was greatly enhanced *in vitro* (Cherepanov et al., 2003). Independently, LEDGF/p75 was identified as HIV-1 IN binding partner in a yeast two-hybrid screen (Emiliani et al., 2005).

The LEDGF mRNA is spliced differently to generate two forms of the protein, LEDGF/p52 and LEDGF/p75. These proteins are differentiated by their C-terminal regions (Ge et al., 1998). Only the p75 form had a direct interaction with IN, while LEDGF/p52, showed no interaction with IN (Cherepanov et al., 2003) and failed to stimulate integration activity *in vitro* (Maertens et al., 2003, Llano et al., 2004). LEDGF/p75 knockdown by siRNA in HeLa and HEK-293T cells led to a decreased accumulation of HIV-1 IN in the nucleus (Maertens et al., 2003). These observations further strengthened the hypothesis that LEDGF/p75 maybe a host factor of HIV-1 integration.

The domain organisation of LEDGF/p75 was determined by limited proteolysis of the recombinant protein, in which a highly conserved domain, resistant to proteolysis, close to the C-terminus of the protein was found (Cherepanov et al., 2004). On deletion of 205 C-terminal residues, the protein failed to interact with IN *in vitro,* and these residues from 326 to 530 were not present in LEDGF/p52. Through further sequence and mutational analysis, it was determined that the 83 amino acids residues 347-429, were highly conserved. This region was minimally sufficient and essential for the interaction with HIV-1 IN. This was identified as the integrase-binding domain (IBD) (Figure 1-13A) (Cherepanov et al., 2004). The structure of IBD was initially determined by NMR, within a LEDGF/p75(347-471) fragment (Cherepanov et al., 2005b). The structure revealed a compact right-handed helical bundle, comprised of four principal α-helices, α1, α2, α4 and α5. Helices α1-α2 and α4-α5 form hairpins and these are connected by α3 (Figure 1-13B). In comparison to the other four helices, α3 is a shorter helix and crosses the helical bundle at its bottom (Cherepanov et al., 2005b).

Amino acid substitutions of LEDGF/p75 residues Ile-365, Asp-366, or Phe-406 disrupted the interaction with HIV-1 IN *in vitro*, while amino acid substitutions of Val-408 and Lys-415 only caused a partial defect in IN interaction (Cherepanov et al., 2005b). Mutational analysis of IN CCD, including Val-165, Arg-166, Leu-172 and Lys-173, caused decreased interaction with LEDGF/p75 *in vitro*. HIV-1 IN residues 165 to 173 were initially suggested to be involved in LEDGF/p75 binding (Cherepanov et al., 2005b).

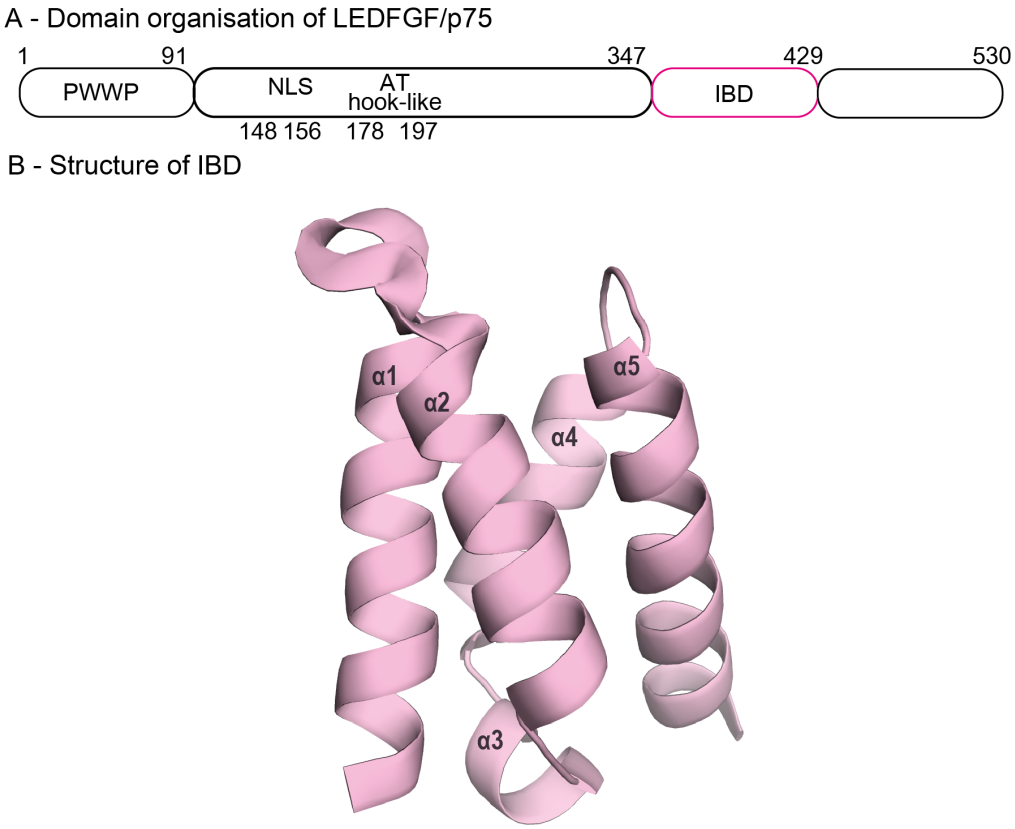

Figure 1-13 LEDGF/p75.

(A) Domain organisation of LEDGF/p75, including the PWWP domain and integrase binding domain (IBD), with the positions of its nuclear localisation signal and AT-hooks indicated; (B) solution structure of IBD (PDB ID 1Z9E), comprising of four α-helices (α1, α2, α4 and α5) arranged as two hairpins, joined by α5 (Cherepanov et al., 2005b).

The structure of HIV-1 IN CCD, with F185K substitution, in complex with LEDGF/p75 IBD was later determined by X-ray crystallography (Cherepanov et al., 2005a). The structure revealed a pair of IBDs, bound at the CCD dimer interface (Figure 1-14A). The two CCD-IBD interfaces are nearly identical, with the complex featuring a near 2-fold symmetry. Within the complex, the CCD formed a dimer, as observed previously (Dyda et al., 1994), and the IBD is nearly identical to its solution structure (Cherepanov et al., 2005b, Cherepanov et al., 2005a) (Figures 1-13B and 1-14A). Thus, neither protein domain significantly changes its conformation upon complex formation.

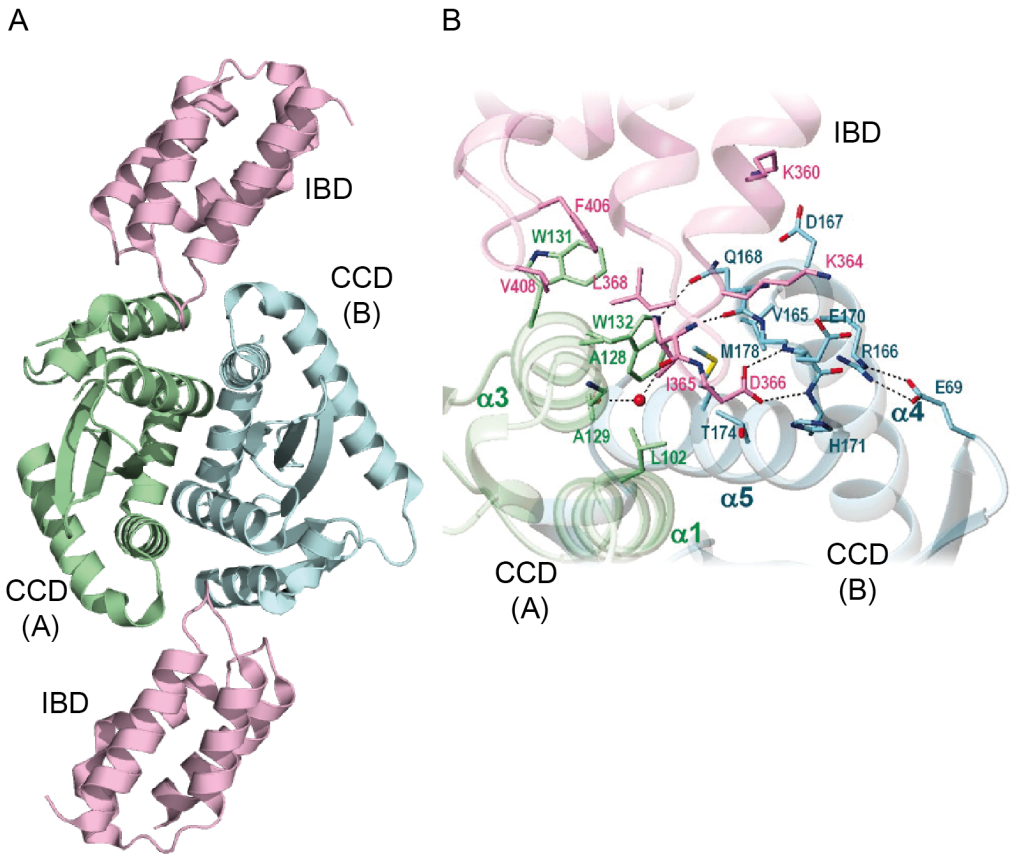

Figure 1-14 Crystal structure of HIV-1 IN CCD in complex with LEDGF/p75 IBD.

(A) A dimer of CCDs (green and blue) in association with two IBDs (pink) (PDB ID 2B4J). (B) Side-chain of LEDGF/p75 Ile-365 extended into the hydrophobic region of IN, with surrounding hydrophobic amino acid residues, Leu-102, Ala-128, Ala-129 and Trp-132 of one subunit and Met-178 and Thr-174 of the second subunit, indicated in sticks and hydrogen bonds are indicated by dashed lines (Cherepanov et al., 2005a, Hare et al., 2009a). Figure (B) was obtained with appropriate permissions (Copyright (2005) National Academy of Sciences).

The LEDGF/p75-binding site is formed by helices α1 and α3 of one IN CCD subunit (indicated in green, Figure 1-14) and α4 and α5, along with the residues connecting the equivalent helices (α4/α5 connector) in the second IN CCD subunit of the dimer (indicated in blue, Figure 1-14). In agreement with the previously reported LEDGF/p75 amino acid residues implicated to play a role in interacting with IN CCD (Cherepanov et al., 2005b), specific contacts by Ile-365, Asp-366, Phe-406 and Val-408 with IN were observed in the crystal structures (Cherepanov et al., 2005a). In particular, the side chain of LEDGF/p75 Ile-365 extends into a hydrophobic region of IN formed by residues Leu-102, Ala-128, Ala-129 and Trp-132 of subunit A and Thr-174 and Met-178 from subunit B (Figure 1.14-B). LEDGF/p75 Phe-406 and Val-408 interact with Trp-131 of IN chain A, while Asp-366 forms hydrogen bonds with IN Glu-170 and His-171 in IN chain B (Figure 1.14-B). Backbone amide of LEDGF/p75 residue Ile-365 makes a hydrogen bond to the main-chain carbonyl group of Gln-168 in IN subunit B (Cherepanov et al., 2005a). These hydrogen bonds play a role in interacting with IN, and when the residues are mutated, LEDGF/p75 fails to bind HIV-1 IN (Cherepanov et al., 2005b).

In spite of apparently lacking a transferable nuclear localisation signal (NLS), HIV-1 IN was localised in nuclei when ectopically expressed in human cells (Pluymers et al., 1999). The cryptic karyophilic property of HIV-1 IN, was later attributed to its interaction with LEDGF/p75, which was lost in cells depleted for LEDGF/p75 (Maertens et al., 2003, Llano et al., 2004). Presence of LEDGF/p75 was also detected in the HIV-1 PICs purified from infected cells and these PICs were active for stand transfer activity *in vitro* (Llano et al., 2004).

LEDGF/p75 belongs to the family of hepatoma derived-growth factor (HDGF) related proteins (HRPs), which also includes HDGF, HRP1, HRP2, HRP3 and HRP4. These proteins are characterised by the presence of N-terminal PWWP domains, containing the Pro-Trp-Trp-Pro motif (Stec et al., 2000, Cherepanov et al., 2004). Upon sequence analysis of the members of this family, it was predicted that HRP2 also contains an IBD (Cherepanov et al., 2004). The protein sequence of human LEDGF/p75 IBD is 48% identical to that of HRP2. Concordantly, the IBD fragment of HRP2 was able to interact with HIV-1 IN *in vitro*, while FLAG-tagged HIV-1 IN from HEK-293T cells readily co-immunoprecipitated both HA-tagged LEDGF/p75 and HRP2 proteins. *In vitro* strand transfer activity of HIV-1 IN was stimulated by HRP2,

in a similar fashion to LEDGF/p75 (Cherepanov et al., 2004). Moreover, both LEDGF/p75 and HRP2 were able to restore integration activity in salt-stripped PICs *in vitro* (Vandegraaff et al., 2006).

The interaction with LEDGF/p75 was detected for a range of lentiviral INs (Cherepanov, 2007). PICs from cells infected with FIV also contained LEDGF/p75 (Llano et al., 2004). BIV, MVV and EIAV INs readily interacted with LEDGF/p75 *in vitro* and this interaction enhanced their strand transfer activities (Cherepanov, 2007). In case of MVV IN, under the conditions tested, *in vitro* strand transfer and intasome assembly was observed only in the presence of LEDGF/p75 (Ballandras-Colas et al., 2017).

1.6 Retroviral integration site selection

The choice of retroviral integration site within host chromosomes is not a random event, and INs display genus-specific features in selecting their integration sites. These features manifest at three vastly different genomic scales: proximal target DNA sequence (~10 bp), local genomic features (*e.g*. relative to locations of genes, promoters, CpG islands), and finally, large chromosomal domains (*e.g*. >1-Mbp sections of the genome characterized by high or low gene density).

1.6.1 DNA sequences immediately surrounding retroviral integration sites

Retroviral INs display a weak preference for the sequence of DNA flanking the integration sites, which was analysed by sequencing genomic DNA from infected cells, as well as *in vitro* integration reactions carried out with genomic or plasmid target DNA. In one of the earliest examples, 29 HIV-1 integration sites were identified, revealing a weak preference for the consensus sequence 5′-GTA(A/T)(T/C)-3′ (Stevens and Griffith, 1996). Subsequently, sequencing of higher number of unique HIV-1 integration sites revealed the preference as 5′-(-3)TDG(int)GTWACCHA(+7)-3′ (nomenclature using standard International Union of Biochemistry mixed base codes), where ‘int’ is the site of integration (generally regarded as position 0), positions of the nucleotides upstream and downstream of this integration site are indicated with ‘-’ and ‘+’ symbols, respectively (Carteau et al., 1998, Holman and Coffin, 2005, Wu et al., 2005). Preferences around

the site of integration for MLV are 5′-(-4)DNST(int)VVTRBSAV(7)-3′ (Holman and Coffin, 2005, Aiyer et al., 2015), while PFV integration preferentially occurs in (-2)WG(int)SNNSCW(+5) (Maertens et al., 2010). A retroviral family specific bias against integrating immediately upstream of a thymidine was reported in HIV-1, MLV, ASLV and PFV (Holman and Coffin, 2005, Wu et al., 2005, Maertens et al., 2010).

Retroviral integration sites display a two-fold symmetry at the integration sites, which arises due to the insertion of both the 3′ ends of vDNA into opposite strands of the target DNA separated by four to six bases (Figure 1-15) (Holman and Coffin, 2005, Ballandras-Colas et al., 2017). This apparent preference for palindromic sequences is a direct consequence of the two-fold symmetry observed in the retroviral intasomes. The variations observed between retroviral integration site preferences as well the target DNA duplication differences is a direct consequence of the intasome-target DNA interactions (Maertens et al., 2010, Aiyer et al., 2015, Serrao et al., 2015).

Target DNA is severely distorted when bound to the intasome, and therefore one important factor is target DNA flexibility. Thus, the more flexible base stacking in pyrimidine-purine combinations (YR) are preferred over the rigid RY stacking (Serrao et al., 2015). The target DNA in PFV intasome was observed to be severely distorted, which is potentially made possible by the integration events preferentially occurring at 5′-(int)YYRR-3′, with the flexible stacking of the YR nucleotides in the 4-bp target site duplication (Maertens et al., 2010). Similarly, the HIV-1 intasome would have access to the YR nucleotides in the 5-bp target site duplication at the 5′-(int)RYNRY-3′, irrespective of the specification of the central nucleotide (Serrao et al., 2015). In comparison to PFV, MLV and HIV-1, alpha- (ASLV), beta- (MMTV) and deltaretroviruses (HTLV-1), displayed less strict preferences for the nucleotide identities present in this region. However, even these displayed a significantly higher preference for the presence of YR nucleotides at the centre of the target site duplications (Serrao et al., 2015).

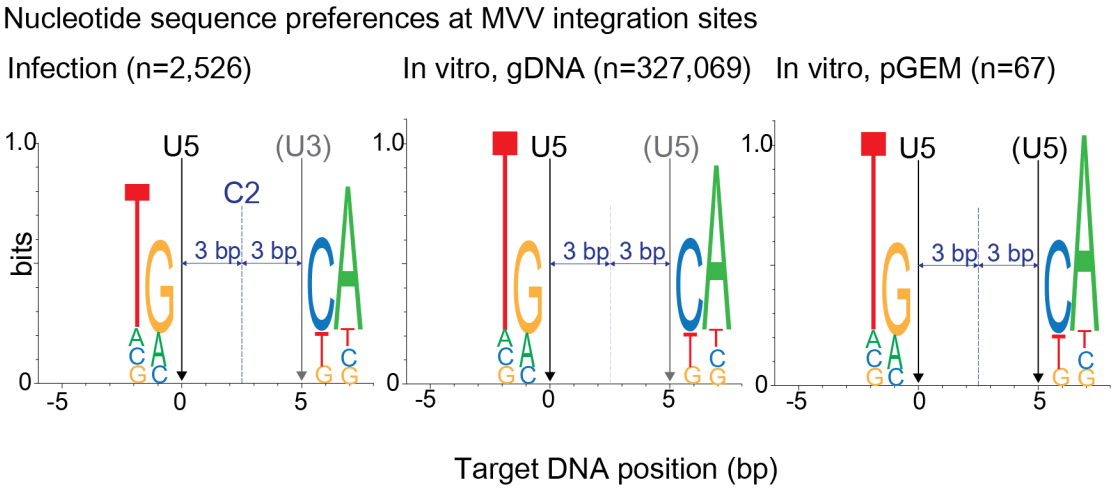

Figure 1-15 Nucleotide sequence preferences for MVV integration.

Preferences of nucleotides flanking the MVV integration sites (shown as sequence logos), with the number of unique sites analysed are indicated (Ballandras-Colas et al., 2017).

1.6.2 Selection of integration site at the level of local genomic features.

Human lymphoid cell lines were infected with HIV-1, from which 524 integration site junctions were cloned and mapped on to the first draft sequence of the human genome (Schröder et al., 2002, Lander et al., 2001). This landmark study demonstrated a clear preference for HIV-1 integration within active transcription units (TUs) and gene-dense regions. Many follow-up studies confirmed these observations (Shun et al., 2007, Singh et al., 2015, Sowd et al., 2016, Li et al., 2020). Following infection with HIV-1 or HIV-derived vectors, a large majority (typically 70-80%) of the integration events occurred in TUs (Mitchell et al., 2004, Ciuffi et al., 2005, Shun et al., 2007). By contrast, random integration is expected to target TUs only 40-50% of the time, without a preference for highly-expressed TUs or gene-dense regions.

The preference for HIV-1 integration within active TUs is now known to depend on the interaction between IN and LEDGF/p75. Thus, knockdown of LEDGF/p75 in Jurkat or HEK-293T cells showed a reduced number of integration events into TUs

and an increase of the integration into chromosomal regions with higher GC content (Ciuffi et al., 2005). Interestingly, the effect on integration site distribution in this study was not accompanied by a defect in integration efficiency, suggesting that the host factor is not essential for HIV-1 infectivity. By contrast, in LEDGF/p75-knockout mouse embryonic fibroblast cells (MEFs), there was a several-fold reduction of HIV-1 infectivity (Shun et al., 2007). A significant proportion of integration events in the knockout cells occurred in unfavourable regions of the genome, including transcriptionally-silent genes. A rescue phenotype was observed when the knockout cells were compensated with ectopically expressed LEDGF/p75, but not LEDGF/p52 (Shun et al., 2007). Integration sites recovered from HIV-1 infected LEDGF/p75-null cells displayed unchanged integration site consensus (Shun et al., 2007). Indeed, while the IN interaction with LEDGF/p75 guides HIV-1 integration into actively transcribing units, the local preference for target DNA sequence is determined by the interactions between IN (as part of the CIC) and target DNA at the site of integration.

Knockdown of HRP2 expression in cells depleted of LEDGF/p75, further reduced HIV-1 integration into TUs (Schrijvers et al., 2012b). LEDGF/p75 and HRP2 were knocked out in MEFs, generating LEDGF/p75-null, HRP2-null as well as double LEDGF/p75- and HRP2-null cell lines (Wang et al., 2012). The viral ability to synthesise vDNA was not affected in these cell lines. Although HIV-1 infectivity and integration in HRP2-null cells were comparable to that in WT cells, strong defects were observed in the cells depleted of both the proteins. These observations indicate that HRP2 compensates for the absence of LEDGF/p75 (Wang et al., 2012). Ectopic expression of LEDGF/p75 and HRP2 in the cells rescued HIV-1 replication and infectivity to WT levels (Schrijvers et al., 2012b, Wang et al., 2012).

A lentiviral genus-specific preference for integration site selection, as well as interaction with LEDGF/p75 was observed. Analysis of SIV integration sites, showed similarities with HIV-1 integration profile. It also displayed a preference for integrating in actively transcribing genes (Crise et al., 2005). Furthermore, non-primate lentiviruses, FIV, BIV and EIAV integration profiles were also determined. Similar to HIV-1 integration, these viruses also displayed a LEDGF/p75-dependent integration targeting in gene bodies (Hacker et al., 2006, Kang et al., 2006, Li et al., 2020).

1.6.3 Other known retroviral IN binding partners for targeting integration

Gammaretroviral integration, as seen in MLV, preferentially occurs in close proximity of transcription start sites (TSSs) and CpG islands (Wu et al., 2003). Unlike HIV-1, MLV has only a weak preference for active genes, and host transcription factors binding in proximity to these genomic features were proposed to be responsible for this preference (Mitchell et al., 2004). Bromodomain and extra-terminal domain (BET) family proteins, were identified by mass spectrometry, with Brd3 and Brd4 as principle binding partners of MLV IN and not for HIV-1 IN (Sharma et al., 2013). The ET domains of Brd2, Brd3 and Brd4 interact with MLV IN *in vitro*, both physically and functionally. Treating cells with bromodomain inhibitors during infection shifted MLV integration away from TSSs and reduced infectivity (Gupta et al., 2013). Co-immunoprecipitation study with different FLAG-tagged MLV IN deletion constructs revealed that the CTD (MLV IN residues 270-409) was necessary and sufficient for the interaction with Brd4. Conversely, only the C-terminal 27 residues of MLV IN (amino acid residues 382 to 409) were sufficient for the interaction with Brd4 ET domain (De Rijck et al., 2013).

HTLV-I displays a modest preference for integration in proximity of TSSs (Melamed et al., 2013). Proteomic analysis of C-terminal FLAG tagged HTLV-I and BLV INs from HEK-293T cells revealed a specific interaction with the B56γ subunit of the PP2A. This interaction was specific for deltaretroviral INs and B56γ colocalised when ectopically co-expressed in human cells. Only the B′ family isoform of the regulatory subunits of PP2A specifically interacted with HTLV-I and BLV IN. *In vitro* strand transfer activities of deltaretroviral INs were greatly stimulated by B56γ (Maertens, 2016). Subsequently, the deltaretroviral intasomes were assembled in complex with B56γ and their structures were determined, elaborated in section 1.5.4 (Barski et al., 2020, Bhatt et al., 2020). However, it remains to be determined, if like the other IN-binding proteins (*e.g.* LEDGF/p75 for HIV-1 and Brd3/4 for MLV) PP2A facilitates integration site selection by HTLV-I.

1.6.4 Additional host-viral protein interactions involved in HIV-1 integration site selection

The HIV-1 CA plays an essential role in nuclear import of the PIC (Yamashita and Emerman, 2004), and one of the facilitators of this process is the interaction of CA with the host factor CPSF6. Disruption of the CA-CPSF6 interaction by amino acid substitutions, in particular by the CA N74D mutation, changed the integration pattern away from gene-dense regions (Koh et al., 2013). CPSF6 is part of the cleavage factor Im (CFIm) (Rüegsegger et al., 1996), which plays a role in alternative polyadenylation process and regulates the length of the 3′ UTRs (Brown and Gilmartin, 2003, Gruber et al., 2012, Martin et al., 2012, Jang et al., 2019). The HIV-1 CA-CPSF6 interaction has also been shown to evade innate sensors in monocyte derived macrophages that cause NFκB and IRF3 activation (Rasaiyaah et al., 2013). When the CA-CPSF6 interaction is abrogated, either by mutating the CA residues interacting with CPSF6 (N74D and P90A) or by depletion of CPSF6, the previously untriggered innate sensors were activated, thereby generating an immune response by the production of type I interferons and supressing infectivity (Rasaiyaah et al., 2013).

In addition to activation of immune response, depletion of CPSF6 in human bone osteosarcoma epithelial cells (U2OS cells) altered the HIV-1 integration site selection pattern (Sowd et al., 2016). Under normal conditions, HIV-1 integration is generally preferred within genes (83%), while a much-reduced preference is observed for CpG islands (4.9% within 2.5 kb) and TSSs (4.1%). In LEDGF/p75-null cells, HIV-1 displayed a significantly reduced integration site targeting in genes (62.8%), with a complementary increase in CpG islands and TSS. Although the abrogation of CPSF6 reduced integration into genes (57%), there was also a decrease in integration in proximity to the CpG islands and TSSs. The CA-N74D mutant also reduced gene-tropic integrations in WT and LEDGF/p75-null cells (Sowd et al., 2016).

Novel HIV-1 integration hotspots were observed in cells depleted of CPSF6 (Sowd et al., 2016). HIV-1 CA N74D mutant also displayed this preference for alternative integration sites in parental cells. In contrast to WT and LEDGF/p75-null cells, a drastically reduced preference for integration near epigenetic markers associated with open chromatin was observed in CPSF6-null cells, with an increased integration

near transcription suppressive H3K9me3 markers (Sowd et al., 2016). Based on the integration profiles observed along the gene bodies, it was inferred that CPSF6 plays a role in ensuring integration within gene-dense regions and steers it away from 3′ ends of the genes. LEDGF/p75 has a more specific role in determining the local position of HIV-1 integration on a gene and targets it towards the 5′ portions of the genes (Sowd et al., 2016).

Strikingly, depletion of CPSF6 led to an increased accumulation of proviral DNA in the peripheral nuclear (PN) zones of the nuclear regions (Achuthan et al., 2018). HIV-1 integration is generally targeted away from lamina-associated domains (LADs) (Marini et al., 2015), which generally contain repressive histone modifications and genes that are transcribed with a reduced frequency (Guelen et al., 2008, Lochs et al., 2019). In the absence of CPSF6, the integration pattern of HIV-1 was shifted towards LADs, with a 40% increase in the number of integration events occurring within 2.5 kb of LADs (Achuthan et al., 2018). The genes that are generally favoured for integration are termed as recurrent integration targeted genes (RIGs). In WT cells, the RIGs were approximately 0.2 Mb in size, that were well expressed and in gene rich areas, away from the LADs. In cells depleted for either CPSF6 or both LEDGF/p75 and CPSF6, the RIGs were generally of larger size (up to 1 Mb), LAD associated and in gene sparse regions, that are less transcriptionally active (Achuthan et al., 2018).

The CA-CPSF6 interaction is of extreme importance in guiding HIV-1 integration towards the euchromatin, which is generally present in the nuclear interior, away from the periphery. Hence, CPSF6 was determined as the ‘master regulator’ to guide HIV-1 integration (Achuthan et al., 2018), while LEDGF/p75 guides integration towards the gene bodies (Sowd et al., 2016). Of note, the CA-CPSF6 interaction is specific for primate lentiviruses and was not observed in other lentiviruses, including FIV and EIAV (Li et al., 2020).

1.7 Integration and the chromatin structure

Integration target site selection is also sensitive to the physiological state changes of DNA, including transcription and distinctive degrees of chromatin condensation. Retroviral integration into nucleosomal DNA was observed to be more efficient than integration into naked DNA, with an increased preference for exposed major groove of the DNA (Pryciak and Varmus, 1992). This was later confirmed when integration efficiency into the outer surface of the bent target DNA was higher than on unbent DNA and it was greatly reduced at the inner face of the bend (Müller and Varmus, 1994). Severely distorted DNA with a wider major groove in the nucleosome was shown to be a preferred target for HIV-1 integration (Pruss et al., 1994). Pyrosequencing of >40,000 unique integration sites from infected cells indicated that HIV-1 integration was preferred towards histone post-translational modifications linked to active transcription, including, H3 acetylation, H4 acetylation, and H3K4 methylation over inhibitory modifications, including H3K27 trimethylation and DNA CpG methylation (Wang et al., 2007).

A similar preference of MLV integration into nucleosomal DNA was observed *in vitro* and *in vivo*, gammaretroviruses also directed their integration into the outward-facing major grooves of DNA (Roth et al., 2011). A differential, genus specific preference for nucleosome packing was observed in retroviruses. PFV and MLV integration is preferred into relatively dense and stable nucleosomes, while HIV-1 and ASLV preferred integrating into poorly dense chromatin regions (Benleulmi et al., 2015).

1.7.1 The interactions of PFV intasome with nucleosomal DNA

As mentioned in section 1.5.3, the target DNA is lodged in the cleft between the symmetric halves of the intasome. Although the intasome does not undergo any structural rearrangement during strand transfer, it induces severe bending of the target DNA duplex, thereby providing the enzyme active sites a direct access to the target scissile phosphodiester bonds (Maertens et al., 2010, Hare et al., 2012). Cryo-EM structure of the PFV intasome in complex with a human nucleosome revealed a previously unknown role for the outer IN subunits and a rather large interacting interface between the intasome and the nucleosome (Maskell et al., 2015). This includes three of the four IN subunits, both the gyres of nucleosomal DNA and one

H2A-H2B heterodimer. Within the intasome, the target DNA binding groove interacts with the nucleosomal DNA, present above one of the H2A-H2B histone heterodimers. The target nucleosomal DNA undergoes a deformation and is lifted ~7 Å from H2A-H2B heterodimer surface. Three loops of the inner IN CCD interact with the C-terminal helix of the H2B and the IN CCD-CCD interface of the intasome allows it to lean against the second gyre of nucleosomal DNA. Residues in the CTD of PFV IN make essential contact with the C-terminus of the histone H2B (Maskell et al., 2015).

More recently, a cryo-EM structure of PFV intasome in complex with the nucleosome was determined after the strand transfer reaction (Wilson et al., 2019). In addition to the previously observed IN CTD interactions with H2B (Maskell et al., 2015), this higher-resolution structure allowed determination of the contacts between the N-terminal H2A tail and the IN CTD. The intasome-nucleosome structures, both before and after strand transfer reaction, revealed comparatively minor distortions within the histone octamer, surrounding the H3-H4 dimer, proximal to the site of integration. However, the nucleosomal DNA arm proximal to the integration site was shifted by 2 bp (Wilson et al., 2019). The structural information obtained from the PFV intasome in complex with the nucleosome further confirms that while the intasome does not undergo structural modifications to perform strand transfer, it is able to considerably distort and slide the nucleosomal DNA to allow entry of both its strands into active sites in optimal conformation for strand transfer.

1.8 IN functions outside of retroviral DNA integration

INs catalyse integration of vDNA, which is an essential step in the life cycle of retroviruses. However, introducing amino acid substitutions in HIV-1 IN often cause defects at other stages of the viral lifecycle, most notably viral particle maturation and reverse transcription (Engelman et al., 1995, Engelman, 1999). Based on their phenotypes, HIV-1 IN mutants are traditionally divided into two groups: (*i*) class I mutants, which are specifically defective for the integration step, and (*ii*) class II mutants, which cause pleiotropic defects in the viral replication cycle, most commonly characterized by a drastic defect in reverse transcription (Engelman, 1999).

While phenotypes of the class I mutants are consistent with IN function, those of class II are more difficult to explain. Thus, viruses carrying mutations in the HIV-1 IN CCD, S81I, P109S, N117K and Y143N were not defective for viral capsid p24 synthesis, envelope gp160 synthesis and processing to gp120 and gp41, and viral packaging, however they were defective for reverse transcription and subsequently integration (Shin et al., 1994, Taddeo et al., 1994). Similarly, mutations within the HHCC motif, changing the HIV-1 IN NTD residues, His-12 and His-16, also caused catastrophic vDNA synthesis defects (Wu et al., 1999). HIV-1 RT and IN were able to interact *in vitro* and the multimerization of IN was not necessary for this interaction. The IN CTD domain was sufficient to interact with RT (Hehl et al., 2004). HIV-1 IN CTD mutants R231E, W243E, G247E, A248E, V250E, and I251E were defective for reverse transcription and integration. These mutants failed to interact with RT *in vitro*, suggesting the importance of IN-RT interaction for efficient vDNA synthesis (Tekeste et al., 2015).

Mature HIV-1 particles contain conical capsids. Most of the viruses carrying IN H12N mutant were either in their immature form or attained aberrant morphological forms. These forms include either the complete absence of RNA or the RNA outside the capsid core (Engelman et al., 1995). HIV-1 particle production was affected when a premature stop codon was inserted in the *pol* gene immediately upstream of the IN coding region (Bukovsky and Göttlinger, 1996).

Recently, cross-linking immunoprecipitation combined with high throughput sequencing (CLIP-Seq) analysis of RNA in the virus indicated that HIV-1 IN displays an increased preference for binding to viral genomic RNA than cellular RNA. Viral RNA complex with IN stabilised the formation of smaller IN oligomers of dimers and tetramers, and ~4 IN units were in complex with 1 molecule of viral RNA (Kessl et al., 2016). HIV-1 IN CTD mutants K264A/K266A and R269A/K273A were defective for binding with viral RNA *in vitro* and CLIP-Seq analysis of viruses with these IN mutations failed to bind to viral RNA. Electron microscopy analysis of these viral particles revealed eccentric cores, with viral RNA outside the capsid core (Kessl et al., 2016). These observations clearly indicate that HIV-1 IN plays diverse roles during its infection. However, the question remains unanswered as to whether this is a lentiviral genus-specific IN function or simply unique to HIV-1.

1.9 HIV-1 inhibitors

Retroviral replication critically depends on unique viral enzymes - RT, protease and IN, which are not present in uninfected human cells. Highly potent and specific HIV-1 inhibitors were designed to block the viral replication by targeting these three viral enzymes. Importantly, in addition to saving lives of HIV-1 infected people, effective antiretroviral therapy (ART) prevents HIV transmission, thereby aiding in making HIV infections fully preventable (WHO, 2019).

Drugs targeting reverse transcription were some of the first clinically approved therapeutics for HIV-1 (Devita et al., 1987). RT inhibitors can be classified into two distinct types, the nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs) (De Clercq, 1995). The NRTIs are nucleoside analogues, which lack the 3′-OH group, and hence block the vDNA synthesis when incorporated into vDNA by the RT (Huang et al., 1990). NNRTIs are drugs targeted to bind at an allosteric single site within a hydrophobic pocket of the p66 subunit of HIV-1 RT (Kohlstaedt et al., 1992) and specifically inhibit HIV-1 replication (De Clercq, 1998, Witvrouw et al., 1999). Binding NNTRIs changes the spatial conformation of the substrate binding site on the RT, thereby blocking its DNA polymerase activity (Spence et al., 1995).

HIV-1 protease inhibitors were developed based on the structure of the enzyme and bind to its active site (Wlodawer and Erickson, 1993, Silva et al., 1996). Based on the finding that the activity of the enzyme depends on the shape of the substrate that fits in the enzyme substrate binding envelope and not the sequence itself (Prabu-Jeyabalan et al., 2002), a substrate-envelope hypothesis was proposed. Subsequent inhibitors were designed based on this hypothesis to fit perfectly in the substrate-envelope of the enzyme active site, in order to prevent the development of drug resistance mutations (Chellappan et al., 2007, Nalam et al., 2010).

HIV-1 IN has been extensively explored as drug target, also due to the absence of a close cellular IN homologue. One of the essential steps in viral integration is the strand transfer, which is generally the target for these drugs. The approved class of IN inhibitors are HIV IN strand transfer inhibitors (INSTIs), which target the active site

of the enzyme. In addition, allosteric IN inhibitors (ALLINIs), also known as LEDGF/p75-IN inhibitors (LEDGINs) are being actively developed.

1.9.1 HIV INSTIs

INSTIs bind to the enzyme active sites only in the presence of vDNA and compete with the target DNA (Hazuda et al., 1999, Espeseth et al., 2000, Pandey et al., 2007). Active sites of retroviral INs are structurally similar, and, as a consequence, INSTIs have been shown to be active at inhibiting strand transfer reactions by a range of retroviral INs (Hare et al., 2010, Barski et al., 2020, Cook et al., 2020), including MVV IN (Valkov et al., 2009, Ballandras-Colas et al., 2017). These small molecules typically harbour co-planar heteroatoms (oxygen and nitrogen), along with a halobenzyl side chain (Figure1-16). WHO recommends the use of Dolutegravir (DTG) in adults, while raltegravir based drug administrations are used as alternative regimes in adults and children (WHO, 2019).

INSTIs were proposed to bind to the catalytic Mg^{2+} cations within the HIV-1 IN active site (Figure1-16) (Grobler et al., 2002). This was shown to be the case when the crystal structures of the INSTIs bound to PFV intasome were determined (Hare et al., 2010). The intasome structures also revealed that the halogenated benzyl groups of the drugs displace the 3′ adenosine of the processed vDNA end away from the IN enzyme active site (Hare et al., 2010, Hare et al., 2011). The presence of the scissile dinucleotide prior to 3′-processing prevents the INSTI engagement, thereby these drugs are specifically inhibiting only the strand transfer step of the integration process (Hare et al., 2012). The intasome structures, hence helped in understanding the roles played by the halobenzyl groups and specific inhibitory effect of the INSTIs.

Development of drug resistant mutants is of great concern with antiviral drugs, including INSTIs. HIV-1 mutations causing resistance to raltegravir inhibition are commonly observed at HIV-1 IN residues Tyr-143, Gln-148, and Asn-155 (Cooper et al., 2008). Although the side chains of residues at positions 148 and 155 are not directly involved interacting with raltegravir, the substitutions at these positions affect the geometry of the enzyme active site, thereby interfering with the stable binding of raltegravir (Hare et al., 2011). The subsequent INSTIs (referred as second-generation INSTIs) were synthesised with an extended metal chelating core, thereby

enabling it to make more extensive contacts within the IN active sites (Figure 1-16B) (Hare et al., 2011, Cook et al., 2020, Passos et al., 2020). The second-generation INSTIs are less susceptible to the development of resistance compared to raltegravir (Hightower et al., 2011). The cryo-EM and crystal structures generated with the INSTIs in complex with the lentiviral intasomes (Cook et al., 2020, Passos et al., 2020) not only provide essential information of designing drugs, but also aid in understanding the mechanisms of drug resistance development (Engelman and Cherepanov, 2021).

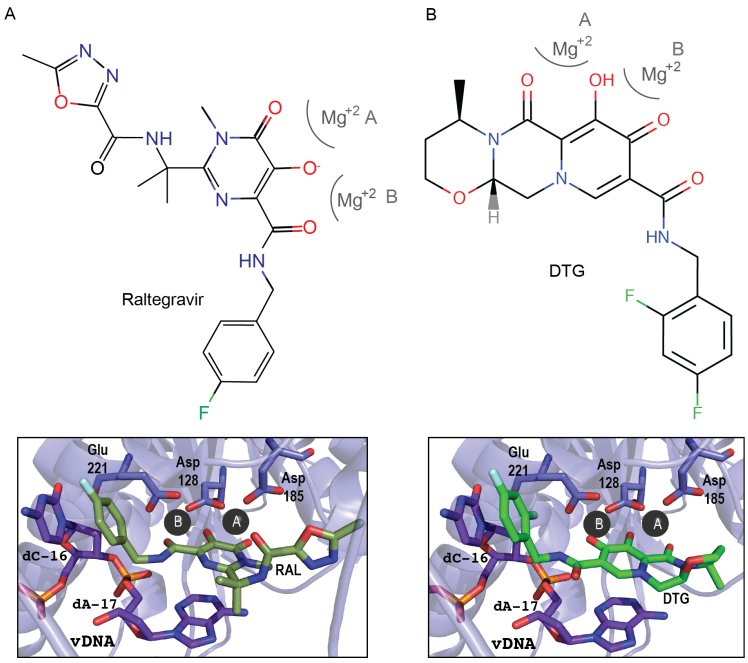

Figure 1-16 INSTIs in complex with the active sites of PFV intasome.

INSTIs harbour the heteroatoms containing metal chelating oxygen atoms, along with a halobenzyl group. The oxygen atoms are oriented towards the metal atoms (A and B) in the PFV enzyme active site – Asp-128, Asp-185 and Glu-221 (indicated in blue sticks). (A) Raltegravir (moss green) and (B) Dolutegravir (DTG, in green) in complex with the active site of PFV intasome. The reactive strand of vDNA (purple) with the 3′-processed end is disengaged from the active site by the inhibitor molecules (Hare et al., 2010 and Hare et al., 2011).

1.9.2 HIV-1 ALLINIs

ALLINIs are small molecule inhibitors of HIV-1 IN, which engage to the LEDGF/p75-binding pocket on the IN CCD, outside of the active site (Christ et al., 2010, Christ et al., 2012, Tsiang et al., 2012). Because this IN region is far less conserved than the active site, ALLINIs are specific to HIV-1 IN only, and inhibit its interaction with LEDGF/p75 (Christ et al., 2010). As can be expected, these compounds can target HIV-1 integration away from genic regions (Feng et al., 2016). However, it was discovered that their main mode of action is in the late stages of HIV-1 replication cycle. These small molecules cause aberrant hyper-multimerization of HIV-1 IN within viral particles. This led to viral particle maturation defects, causing eccentric phenotype, much like that of class II HIV-1 IN mutants, with genomic RNA mislocated outside the capsid core (Jurado et al., 2013). Although ALLINIs bind to the IN CCD similar to LEDGF/p75, they cause different types of IN multimerization. While LEDGF/p75 promotes functional tetramerization of HIV-1 IN (McKee et al., 2008, Hare et al., 2009a), ALLINIs cause protein–protein interactions involving formation of an unnatural CTD-CCD interface (Feng et al., 2016, Gupta et al., 2016, Gupta et al., 2021). These compounds are not yet in clinical use, and require further development.

1.10 Research objectives for this study

Lentiviral integration has been studied extensively and recently, the structures of different lentiviral intasomes (MVV, HIV-1 and SIV_{rcm}) were determined as the largest intasome assemblies *in vitro* (Ballandras-Colas et al., 2017, Passos et al., 2017, Cook et al., 2020). These were at least 2- to 4-fold larger than the previously observed spumaviral, delta-, alpha- and betaretroviral intasome assemblies (Hare et al., 2010, Barski et al., 2020, Bhatt et al., 2020, Ballandras-Colas et al., 2016, Yin et al., 2016, Pandey et al., 2021). In contrast to HIV-1, MVV intasome was assembled using the WT form of the recombinant IN protein (Ballandras-Colas et al., 2017, Passos et al., 2017). The cryo-EM 2D and 3D classes of MVV intasome particles, unlike SIV_{rcm}, were monodispersed and formed individual units of hexadecameric assemblies (Ballandras-Colas et al., 2017, Cook et al., 2020). Complete structural

information of the lentiviral intasome was only obtained with MVV, thereby it appears to represent an ideal system to study lentiviral integration *in vitro*.

The important question at the start of this project was whether this expanded intasome architecture was absolutely necessary for the activity of the lentiviral INs and if it was possible to assemble a functional lentiviral intasome of a smaller size (*e.g.* a tetrameric or an octameric assembly).

Initially, the importance of this assembly was tested by introducing amino acid substitutions targeting the unique features of the intasome. These proteins were characterised biochemically *in vitro*; their ability to assemble into intasomes and perform strand transfer reactions was tested. A selected set of mutants were also tested for infectivity in the context of MVV single cycle virus system. Characterisation of the mutants revealed that the CTD-CTD interface interactions, as well the α-helical confirmation of the CCD-CTD linker are important for the intasome assembly and integration.

Similar to the other tested lentiviral integration patterns, MVV was also shown to preferentially integrate into actively transcribing genes (Ballandras-Colas et al., 2017), however its dependence on the host factor LEDGF/p75 for integration targeting was unknown. On the other hand, MVV intasome assembled *in vitro* only in the presence of LEDGF/p75 (Ballandras-Colas et al., 2017). Hence, the importance of this host protein was tested during MVV infection in both human and sheep cell lines knocked out for LEDGF/p75. Similar to the previously analysed lentiviruses, MVV integration into TUs is dependent on LEDGF/p75, however the presence of the host protein is not essential for MVV infectivity.

Chapter 2. Materials & Methods

2.1 General materials

2.1.1 Common stocks and growth media

A list of common stock solutions and growth media are listed in Table 2-1 and Table 2-2.

Table 2-1 Stocks of solutions and their storage

Stock solutions	Details
Ampicillin	120 mg/ml, stored at -20°C
Isopropyl β- d-1-thiogalactopyranoside (IPTG)	100 ug/ul, stored at -20°C
Bis-tris propane (BTP) pH 7.0	1 M, stored at room temperature
Phenylmethylsulfonyl fluoride (PMSF)	100 mM, stored at -20°C
Imidazole	2 M, stored at 4°C
Dithiothreitol (DTT)	1 M, stored at -20°C
50x TAE	2 M Tris Base, 1 M Acetic acid and 100 mM Ethylenediaminetetraacetic acid (EDTA), stored at room temperature
Branched PEI	10% PEI in Milli-Q water, pH 7.0, stored at -20°C
4x Laemmeli buffer	8% w/v SDS, 40% v/v glycerol, 0.02% w/v bromophenol blue, 10% w/v DTT, 240 mM Tris-HCl pH 6.8, stored at -20°C
10x Tris-glycine-SDS (TGS) buffer	250 mM Tris base, 2 M glycine, 0.01% w/v SDS, stored at room temperature
2x lysis buffer	0.25% Triton X-100, 50 mM KCl, 100 mM TrisHCl, pH7.4, 40% glycerol, stored at room temperature for upto a month
10x core buffer	50 mM $(NH_4)2SO_4$, 200 mM KCl and 200 mM Tris–HCl pH 8.3, stored at room temperature for upto six months, diluted to 1x concentration immediately before use
5x passive lysis buffer (PLB)	125 mM Tris-HCl pH 7.8, 10 mM DTT, 10 mM EDTA, 50% glycerol and 5% Triton X-100, stored at room temperature
Luciferase assay reagent (LAR)	1 mM D-luciferin salt, 3 mM ATP salt, 15 mM $MgSO_4.7H_2O$, 30 mM Hepes pH 7.8, aliquoted and stored at -80°C

Table 2-2 Growth media used in this study

Growth media	
SOC broth	2% Tryptone, 0.5% Yeast extract, 10 mM NaCl, 2.5 mM KCl, 20 mM $MgCl_2$, 20 mM $MgSO_4$, 20 mM D-Glucose
Luria-Bertani broth (LB)	1% w/v tryptone, 0.5% w/v yeast extract, 1% w/v NaCl

2.1.2 Bacterial strains

Two types of *E.coli* bacterial strains were used in this study and their genotypes are listed in Table 2.3. The DH5α strain (New England Biolabs) was used for plasmid amplification. Mutations in its *endA1* and *recA1* genes prevent degradation of plasmid DNA, as well as inhibit homologous recombination to aid in clone stability (Taylor et al., 1993, Roca and Cox, 1997). The PC2 strain used for recombinant protein expression, was derived from a spontaneous T1 phage-resistant mutant of PC1 cells (Cherepanov, 2007). This strain's origin was from BL21(DE3)pLysS cells that were transduced with a P1 phage carrying a deletion mutant of the *E. coli endA* gene in combination with a tetracycline resistance cassette (Cherepanov and Wackernagel, 1995, Cherepanov et al., 1999). *E. coli* BL21 cells naturally lack Ion protease and the ompT outer membrane protease. The *endA* gene encodes for Endonuclease-1, the major non-specific periplasmic nuclease that degrades DNA (Lin, 1992). Because Endonuclease-1 can contaminate recombinant protein preparations, it is advantageous to use *endA*-deleted strains for protein production, especially when the purified protein's activity is assessed in the presence of plasmid DNA. The lack of these nucleases and proteases made it an ideal system to transform plasmids and express proteins. The plasmid pLysS constitutively expresses phage T7 lysozyme at low levels, which inhibits the T7 RNA polymerase, reducing the basal expression of toxic protein products prior to induction and minimizing the chances of plasmid loss. The expression is induced by the addition of isopropyl β-d-1-thiogalactopyranoside (IPTG), which then allows the expression of desired protein in the cells (Studier, 1991).

Table 2-3 Genotype details of bacterial strains used

Bacterial Strain	Genotype
DH5α	F– φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1
PC2	BL21(DE3), endA::TetR, T1R, pLysS; Endonuclease A-deficient strain, recovered as a spontaneous T1 phage-resistant mutant of PC1

2.2 General methods for plasmid DNA construction

2.2.1 Polymerase chain reaction (PCR) for DNA amplification

PCR reactions were carried out using PfuUltra II fusion hot start DNA polymerase (Agilent). This chimeric thermostable enzyme represents a fusion of *Pyrococcus furiosus* DNA polymerase (Pfu) with *Sulfolobus solfataricus* Sso7d protein, which greatly enhances processivity of DNA synthesis (Wang et al., 2004). A 100-µl reaction mixture contained 30-250 ng template DNA, 0.25 mM each deoxynucleotide triphosphate (dNTP) [deoxyribo-adenosine triphosphate (dATP), deoxyribo-thymidine triphosphate (dTTP), deoxyribo-cytosine triphosphate (dCTP) and deoxyribo-guanosine triphosphate (dGTP)], 0.2 µM each oligonucleotide primer, with 1 µl of enzyme in PfuUltra II reaction buffer, supplied with the enzyme as a 10x concentrate. All PCR reactions were carried out using Thermocycler C1000 (Bio-Rad). The PCR reactions were carried out with appropriate annealing temperatures and extension times, based on the primers and the PCR product size, respectively. The resulting PCR products were separated by agarose gel electrophoresis through 1% – 1.5% w/v agarose gels in TAE buffer (40 mM Tris acetate and 1 mM EDTA at pH 8.3), diluted from a 50x stock solution (Table 2-1) and supplemented with 1 µg/ml Ethidium Bromide for DNA visualisation.

In some instances, PCR was carried out using *Thermus aquaticus* DNA polymerase (Taq). This enzyme does not possess a proofreading 3′-5′ exonuclease activity, but can be more robust than Pfu fusion. Moreover, the resultant PCR products are extended by a single dA at either 3′ end, allowing TA cloning into specialysed vectors (see section 2.3.3). A 100-µl reaction mixture contained 30 - 250 ng template DNA, 2 mM $MgSO_4$, 0.25 mM each dNTP, 0.2 µM each oligonucleotide primer, with 2 µl

enzyme (Life Technologies) in Taq polymerase reaction buffer, supplied with the enzyme as 10x concentrate. The PCR reactions were carried out as described above.

2.2.2 Overlap extension PCR

Overlap extension PCR method was used to introduce mutations in the CTD of MVV IN and the CCD-CTD linker of MVV IN, as well as to construct the template for generation of standard curves in late reverse transcription (LRT) quantitative PCR (qPCR) assays. In the first round of PCR, the thermocycler parameters were: 95ºC for 30 sec (initial denaturation), followed by 25 cycles of 95ºC for 30 sec (denaturation), 50ºC for 60 sec (annealing), 72ºC for 2 min (extension); a final extension step was carried out at 72ºC for 5 min. Two 100-µl reactions were performed in parallel that contained 250 ng template DNA (for example, pCPH6P-MVV-IN encoding WT MVV IN), 0.25 mM each dNTP, 0.2 µM each primer (outer primers PC1095 or PC1096 and appropriate inner primers at the site of mutation), 1 µl enzyme in Pfu Fusion reaction buffer. In the second round of PCR, approximately equimolar amounts of overlapping PCR products from the first round PCR were used as templates and amplified with 0.2 µM outer primers PC1095 and PC1096. The oligos and their sequences used to introduce mutations are listed in Appendix Table 8.3. Likewise, pCVW-CG-Luc plasmid was used as template DNA for the synthesis of the plasmid to generate standard curve for LRT qPCR, using outer primers PC1803 or PC1806 and inner primers PC1804 or PC1805 (Appendix Table 8.3).

2.2.3 DNA purification by agarose gel electrophoresis

PCR products were separated by electrophoresis in 1% agarose gels in TAE buffer (diluted from a stock of 50x, Table 2-1) and were purified from gel slices using QIAquick gel extraction kit (Qiagen), as per the manufacturer's instructions. Briefly, slices of agarose gel containing DNA fragments of interest were dissolved in 3 volumes of QG buffer (supplied) by incubation at 50ºC for 10 min. In the presence of high concentrations of salt in QG buffer, only the DNA is bound to the silica gel membrane filter in a QIAquick spin column. The impurities, including agarose and electrophoresis buffer components pass through the column and are discarded. DNA

retained in the spin columns was washed with Buffer PE (supplied) and eluted in 30 - 50 µl Milli-Q water.

2.2.4 Restriction digestion of DNA

1 to 2 µg purified PCR product or 5 µg plasmid DNA was combined with 10 to 20 U (1 - 2 µl) each relevant restriction enzyme (New England Biolabs) in the presence of recommended buffer (usually CutSmart Buffer, New England Biolabs). Reactions were made up to a final volume of 50 or 100 µl with Milli-Q water and incubated at 37ºC for 6-12 h. The restriction enzymes used for plasmids and PCR products are listed in Appendix Table 8.1. Restriction enzyme-digested PCR products or plasmids were separated by electrophoresis through 1% (w/v) agarose gels in TAE buffer and purified using QIAquick gel extraction kit (Qiagen).

2.2.5 DNA ligation

A 3:1 molar ratio of insert DNA fragment (typically purified PCR product, digested with appropriate restriction enzymes) and restriction enzyme-digested plasmid were ligated using 1 Weiss U (1 µl) T4 phage DNA ligase (Life Technologies) in reaction buffer (50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 1 mM ATP, 1 mM dithiothreitol (DTT) and 5% w/v polyethylene glycol-8000, supplied with the enzyme as a 5x concentrate) in a final reaction volume of 10 µl. Reactions were incubated for at least 1 h, but typically overnight, at 16˚C to allow ligation of DNA fragments.

2.2.6 Transformation of competent *E. coli* cells

1 to 5 µl DNA ligation reaction or 250 ng plasmid DNA was added to 50 µl chemically competent DH5α strain of *E. coli* cells (New England Biolabs) and incubated on ice for 20 min. The bacteria were then incubated at 42ºC for 45 sec, and supplemented with 200 µl SOC medium. Transformed bacteria were shaken gently at 37ºC for 1 h before plating on selective LB agar Petri dishes containing appropriate antibiotic(s) (typically 120 µg/ml ampicillin). The dishes were incubated overnight at 37ºC, and single colonies were used to inoculate 5-ml liquid LB broth cultures supplemented with appropriate antibiotic(s). The cultures were grown overnight with vigorous shaking at 37ºC.

2.2.7 Small-scale preparation of plasmid DNA (miniprep)

Plasmid DNA was purified from 5-ml cultures using commercial QIAprep Spin Miniprep protocol (Qiagen). Bacterial pellet was resuspended in buffer P1, supplemented with RNase A (supplied with the kit) and lysed in buffer P2 (supplied). P2 contains high concentration of sodium hydroxide that dissolves cells and denatures protein and nucleic acids. The lysate was neutralised by buffer N3 (supplied), which causes co-precipitation of cellular debris and denatured chromosomal DNA. By contrast, intact plasmid DNA renatures quickly and stays in solution. The lysate was centrifugated at 10,000 *g* for 20 min to pre-clear, and supernatant was applied on to the silica membrane in the QIAprep spin column (supplied). Plasmid DNA was bound to silica membrane in the spin columns. Following washes with ethanol-containing PE buffer (supplied), pure plasmid DNA was eluted with 30 - 50 µl Milli-Q water.

2.2.8 Large-scale preparation of plasmid DNA (maxiprep)

To isolate pure and highly concentrated plasmid DNA for the use in transfections, the Plasmid Maxi kit (Qiagen) was used as per the manufacturer's instructions. A 3-ml LB broth culture inoculated with a single bacterial colony and grown for 5 - 8 h at 37ºC was used to inoculate 200 - 300 ml LB broth culture supplemented with the relevant antibiotic. This was grown overnight at 37°C with vigorous shaking at 250 rpm. Bacterial culture pelleted by centrifugation at 6,000 rpm (JA-10 rotor, Beckman Coulter) for 10 min was resuspended in P1 buffer complemented with RNase A (supplied) and lysed by addition of 10 ml P2 buffer (supplied). The lysate was then neutralised with 10 ml chilled P3 buffer (supplied) and centrifuged at 20,000 *g* for 30 min at 4°C to separate the precipitated contaminants. The clear supernatant, containing plasmid DNA, was loaded on Plasmid Maxi column, equilibrated with QBT buffer (supplied). Impurities were washed away with QC buffer (supplied), while plasmid DNA binds to the anion exchange resin under low salt conditions. Plasmid DNA was then eluted with the high-salt QF buffer (supplied) and precipitated with 25% isopropanol by centrifugation at 4,000 rpm (Rotor S-4-104, Eppendorf) at 4°C for 1 h. Plasmid DNA was washed with 70% ethanol by centrifugation at 4,000 rpm (Rotor S-4-104, Eppendorf) for 30 min at room temperature. The DNA pellet was

dissolved in 0.5 ml Milli-Q water. Purified DNA was quantified by spectrophotometry at 260 nm using a Nanodrop instrument (Thermo Scientific) and stored at -20°C for further use.

2.2.9 Sanger DNA sequencing

The plasmids, or in some instances PCR products, were sequenced using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher). DNA was purified from overnight 5-ml LB cultures inoculated with single bacterial colonies (section 2.2.7). The reaction mixture contained 185 ng purified plasmid DNA, 3.2 pmol oligonucleotide primer (*e.g*. T7 T primer, listed in Appendix Table 8.3), 8 µl Big Dye master mix (diluted 1:16 with 2.5x Big Dye dilution buffer, supplied as 5x concentrate) and Milli-Q water in a total volume of 20 µl. The sequencing reaction was carried out with the reaction parameters: 96ºC for 1 min (ramp rate of 2.5º/sec) for 1 cycle, followed by 96ºC for 10 sec, 50ºC for 5 sec and 60ºC for 4 min (ramp rate of 1º/sec) for 25 cycles.

Extension products were purified from unreacted BigDye terminators and other reaction components using DyeEx 2.0 spin columns (Qiagen) following manufacturer's instructions. First, each spin column was pre-treated by brief vortexing and centrifugation at 750 *g* for 3 min. Reaction samples were applied to the resin in the spin column and centrifuged at 750 *g* for 3 min. Samples were eluted into fresh 1.5-ml Eppendorf tubes and dried in Vacufuge Plus (Eppendorf) centrifugal evaporator. The samples were then submitted to the Genomics Equipment STP (The Francis Crick Institute) for Sanger Sequencing. The sequencing data, made available in '.ab1' file formats, were analysed using ContigExpress software (Vector NTI, Thermo Fisher).

2.3 DNA constructs

The plasmids generated and used in this study (Appendix Table 8.1) were verified by Sanger sequencing at The Genomics Equipment STP (The Francis Crick Institute). References and sources of the plasmids available commercially or provided by collaborators were indicated in Appendix Table 8.1.

2.3.1 DNA constructs for bacterial expression of MVV IN and LEDGF/p75

The constructs used for expression of WT MVV IN and human LEDGF/p75 were provided by Prof. Peter Cherepanov. Amino acid sequence of MVV KV1772 IN is available from UniProt database (www.uniprot.org) under accession number P35956 (positions 1226 - 1506). Recombinant IN proteins were expressed using the pCPH6P-MVV-IN plasmid, illustrated in Figure 2-1 (Cherepanov, 2007). Amino acid sequence of human LEDGF/p75 sequence corresponds to UniProt entry O75475. Recombinant LEDGF/p75 was expressed using the pCPH6P-LEDGF plasmid (Cherepanov, 2007). Extinction coefficients for the proteins at 280 nm were calculated using the protein sequence on the ExPASy ProtParam (https://web.expasy.org/protparam/). Both MVV IN and LEDGF/p75 were present between the *Xma*I and *Xho*I restriction sites on pCPH6P bacterial expression plasmid, which encodes a recognition sequence for the human Rhinovirus type 14 3C (HRV14 3C) protease downstream of the hexahistidine (His_6) tag (Figure 2-1).

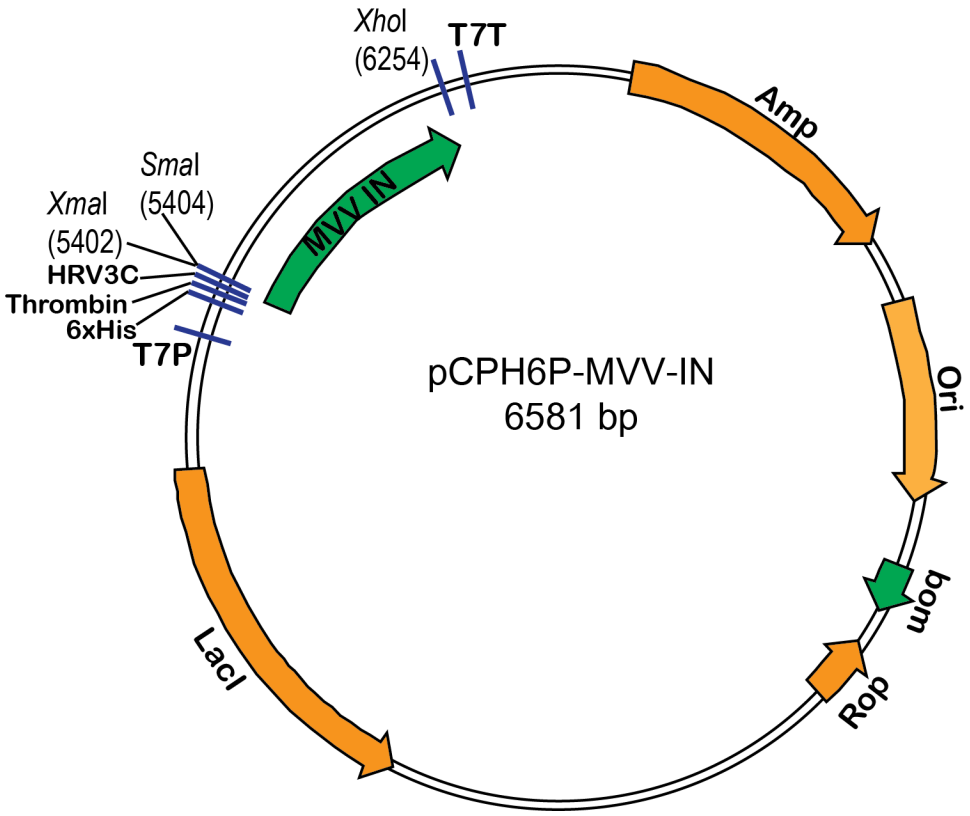

Figure 2-1 pCPH6P plasmid used for MVV IN expression.

pCPH6P-MVV-IN plasmid was used for the expression and purification of MVV IN proteins. It contains a T7 Lac promoter, sequences for His6-tagged protein and ampicillin resistance. Figure adapted from Vector NTI with appropriate permissions.

2.3.2 Retroviral constructs for generation of single-cycle viruses

The production of single cycle viruses required co-transfection of HEK-293T cells with a combination of plasmids. These generally included a packaging construct (encoding viral Gag-Pol polyprotein), a reporter transfer vector (encoding either green fluorescence protein (GFP) or firefly luciferase), and an envelope construct (encoding vesicular stomatitis virus glycoprotein G (VSV-G)).

2.3.2.1 ***Plasmids for the generation of MVV single-cycle virus***

The MVV single cycle virus vector system was designed and supplied by Dr. David Griffiths' laboratory (Moredun Research Institute, UK). It comprises four plasmids (Figure 2-2): MVV Gag-Pol packaging construct pCAG-MV-Gagpol-CTEx2, reporter plasmid pCVW-CG-GFP, envelope construct pMD2.G, and a codon-optimised MVV Rev expression construct pCMV-VMV-Rev.

The packaging construct (pCAG-MV-Gagpol-CTEx2) contains chicken beta-actin/CMV promoter/enhancer element (CAGP), MVV *gag-pol* gene, Mason-Pfizer monkey virus (MPMV) constitutive transport element (CTE) and SV40 late polyadenylation signal. The reporter plasmid (pCVW-CG-GFP) contains a human cytomegalovirus immediate early promoter (CMVP), repeat region of MVV LTR (R), unique 5′ region of MVV genome (U5), unique 3′ region of MVV (U3) LTR, putative encapsidation element (ψ), partially deleted region of MVV *gag* gene (Δ*gag*), MVV Rev-responsive element (RRE), central polypurine tract and central termination sequence (cPPT/cts), *gfp* gene, woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) and MVV LTR, including U3, R and U5. The codon-optimised Rev-expression plasmid (pCMV-VMV-Rev) harbours MVV *rev* gene under the control of CMVP.

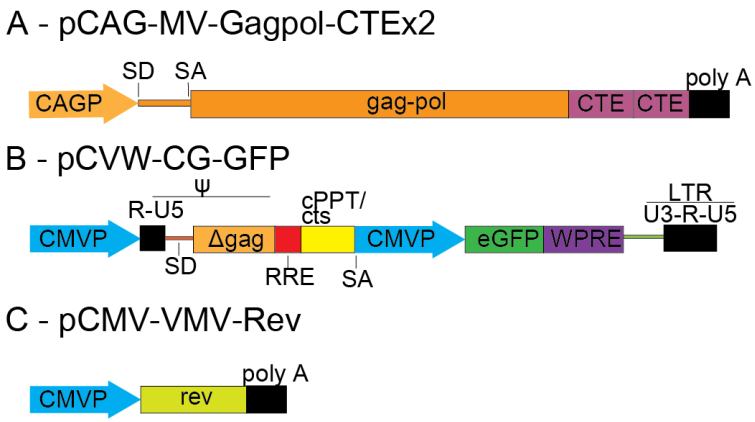

Figure 2-2 MVV single-cycle viral vector system.

The MVV single cycle plasmid system consists of (A) pCAG-MV-Gagpol-CTEx2 (packaging construct), (B) pCVW-CG-GFP (reporter plasmid), (C) pCMV-VMV-Rev (codon-optimised Rev plasmid) and pMD2.G (not depicted in the figure). The MVV packaging construct contains a chicken beta-actin/CMV promoter/enhancer element (CAGP), MVV *gag-pol* gene, Mason-Pfizer monkey virus (MPMV) constitutive transport element (CTE) and SV40 late polyadenylation signal (polyA signal). The reporter plasmid and the codon optimised rev plasmid are under the control of human cytomegalovirus immediate early promoter (CMVP). In addition, the reporter plasmid contains the MVV LTR regions (U3-R-U5), putative encapsidation element (ψ), partially deleted region of MVV *gag* gene (Δgag), MVV Rev-responsive element (RRE), central polypurine tract and central termination sequence (cPPT/cts), *gfp* gene and woodchuck hepatitis virus post-transcriptional regulatory element (WPRE). The codon optimised Rev plasmid harbours the MVV *rev* gene with the polyA signal.

In a few instances, modifications were made to the MVV plasmid system obtained. Restriction sites, *Age*I and *Xho*I, flanking IN were introduced by creating silent mutations in pCAG-MV-GagPol-CTE2x, thereby allowing for the replacement of the WT IN coding region with the mutations, detailed in section 4.2. The 2.5 kb region between *Pas*I and *Dra*III regions on the plasmid was initially amplified by PCR using primers PC1246 and PC1247 (Appendix Table 8.3). The PCR products were ligated into restriction digested pBluescript SK(-) to generate pBS-MVVIN-CTE2x.

Complimentary PCRs were carried out using pBS-MVVIN-CTE2x as template DNA to first introduce *Xho*I restriction site downstream of the IN coding region, with the primers PC1254 and PC1255. After the successful insertion of *Xho*I restriction site, silent mutations were introduced immediately upstream of IN to introduce *Age*I, using primers of PC1252 and PC1253 to generate pBS-MVVIN-*Age*I/*Xho*I-CTE2x. The modified 2.5 kb fragment was introduced into the original pCAG-MV-GagPol-CTE2x, to generate pCAG-MV-GagPol-CTEx2-IN-Age/XhoI. WT-KV1772 and mutant INs were amplified from the previously generate protein expression plasmids (*e.g.* pCPH6P-MVV-IN), using primers PC1263 and T7-terminator. Restriction digested PCR fragments were introduced into AgeI and XhoI restriction digested pCAG-MV-GagPol-CTEx2-IN-Age/XhoI, to generate pCAG-MV-GagPol-CTEx2-IN.KV1772-Age/XhoI. The plasmids produced at each stage were verified by Sanger sequencing, before proceeding to the next steps.

Furthermore, *gfp* gene in pCVW-CG-GFP was replaced with a DNA fragment encoding firefly luciferase and the newly formed plasmid was referred to as pCVW-CG-Luc, detailed in section 4.3.

2.3.2.2 ***Plasmids for the generation of MLV vector***

The MLV retroviral vector system (Ulm et al., 2007), available in the laboratory, was used for expression of *Ovis aries* LEDGF/p75 (*Oa*LEDGF/p75) in HEK-293T LHKO cells. The system comprises pCG-GagPol (an MLV packaging construct) and the envelope construct pCG-VSVG.

The pCG-GagPol plasmid contains of a CMV promoter and a β-globin intron, along with a β-globin polyadenylation signal for the expression of viral sequences (Ulm et al., 2007). The pQCXIP vector (Clontech) allows selection of stably transduced cells with puromycin and contains multiple cloning sites, including *Not*I*, Mfe*I*, Age*I*, Xho*I *and Bam*HI. A DNA fragment encoding WT or mutant *Oa*LEDGF was ligated between *Not*I and *Bam*HI sites of pQCXIP.

2.3.3 Plasmid to generate standard curves for LRT qPCR

The plasmid used to generate standard curve to quantify viral DNA post-infection was synthesised by overlapping PCR, described in section 2.2.2, using pCVW-CG-Luc plasmid as template DNA. The PCR product was ligated into pCR4-TOPO (Invitrogen) to generate pCR4-TOPO-MVV-LRT plasmid; the sequence of the construct was confirmed by Sanger sequencing analysis.

2.3.4 Plasmids encoding CRISPR/Cas9

The sgRNA oligos (F′ and R′) targeting *Oa*LEDGF/p75 IBD were annealed in a total reaction volume of 10 µl, containing 10 µM each of the oligos, 10 U T4 Polynucleotide Kinase (T4 PNK, Thermo Scientific) in T4 ligation buffer (supplied with the enzyme as a 10x concentrate). The reaction was carried out at 37ºC for 30 min, 95ºC for 5 min and cooled down to 25ºC at the rate of 5ºC/min. The annealed oligos were stored at -20ºC for future use.

The CRISPR/Cas9 encoding plasmid, pSpCas9(BB)-2A-GFP was digested with the restriction enzyme *Bbs*I. The reaction mixture contained 3 µg template DNA, 1 µl (20 U) *Bbs*I-HF (New England Biolabs) in CutSmart buffer (supplied with the enzyme as a 10x concentrate) in a total reaction volume of 100 µl. Digested DNA was separated by agarose gel electrophoresis and purified using QIAquick gel extraction kit (Qiagen), as described in section 2.2.3. 100 ng purified linearized plasmid was ligated with 2 µl annealed oligo duplex (diluted from 50 µM stock to a final concentration of 0.5 µM or 0.25 µM in Milli-Q water prior to use) in the presence of 1 Weiss U T4 DNA ligase (Life Technologies) in T4 DNA ligase reaction buffer (supplied as a 5x concentrate with the enzyme) in a final reaction volume of 20 µl. The resulting plasmids pSpCas9(BB)-2A-GFP-sgRNA1, pSpCas9(BB)-2A-GFP-sgRNA2, and pSpCas9(BB)-2A-GFP-sgRNA3 were verified by sequencing analysis.

2.4 Tissue culture

Cell lines generated and used in this study are listed in Appendix Table 8.2. All cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco, Life Technologies) supplemented with 10% foetal bovine serum (FBS; obtained from the Cell Services STP, The Francis Crick Institute) and antibiotic-antimycotic cocktail (supplied as a 100x concentrate, Life Technologies), henceforth referred as complete medium. Cells were cultured in humidified, 5% CO_2 atmosphere at 37ºC.

2.4.1 Transfection of HEK-293T cells

Transfection of HEK-293T cells (obtained from The Cell Services STP, The Francis Crick Institute) to generate MVV single-cycle virus particles was performed using 0.1% polyethylenimine (PEI, Sigma-Aldrich, catalog #408727). PEI was diluted from 10% stock (w/v, made in water) with 100 volumes of PBS immediately before use). $10x10^6$ cells were seeded in complete medium 16 - 18 h prior to transfection in 15-cm culture dishes (Falcon, Fisher Scientific). 58.5 µl 0.1% PEI was added to pre-warmed 2.5 ml Opti-MEM reduced serum medium (Life Technologies). Plasmid DNA was then added in the ratios specified in Table-2.4. The transfection mix was incubated at room temperature for 15 - 20 min. In the meantime, complete medium was replaced with pre-warmed Opti-MEM on the seeded cells. The transfection mix was added drop-wise on to the cells, without disturbing the established monolayer. 18 h post-transfection, the medium was replaced with complete medium. 48 - 72 h post-transfection, conditioned medium containing viral particles was harvested. Cell debris were pelleted by centrifugation at 200 *g* for 10 min at room temperature, and supernatant was filtered through a 0.45-µm filter and transferred to 35 ml UltraClear ultracentrifuge tubes (Thermo Fischer). The viral supernatant was centrifuged at 30,000 rpm for 1.5 h at 25ºC in Optima XPN Ultracentrifuge (Beckman Coulter) with an SW 32 Ti swinging bucket rotor (Beckman Coulter). Following ultracentrifugation, the supernatant was gently removed by aspiration and the viral pellet was resuspended in 300 - 500 µl complete medium. Concentrated virus preparations were stored at -80ºC in small aliquots.

Table 2-4 Ratio of the plasmids used for production of single cycle viral vectors

No.	Virus	Plasmid	µg of DNA	Total DNA
1	MVV	pCAG-MV-Gagpol-CTEx2	10.5	35 µg
		pCMV-VMV-Rev	3.5	
		pMD2.G	5.25	
		pCVW-CG-Luc	15.75	
2	WT or D366N OaLEDGF MLV	pCG-GagPol	2	5 µg
		pCG-VSVG	0.5	
		pQFLAG-Puro-*Oa*LEDGF/p75-WT/D366N	2.5	

Transfection of HEK-293T cells to generate MLV single cycle viral particles with *Oa*LEDGF/p75 was performed using Xtreme Gene 9 transfection reagent (Sigma-Aldrich). $3.5x10^6$ cells were seeded in complete medium 16 - 18 h prior to transfection in 10-cm culture dishes. 15 µl Xtreme Gene 9 was first added to 500 µl pre-warmed Opti-MEM. Plasmid DNA was then added in the amounts specified in Table 2.4. The transfection mix was incubated at room temperature for 15 - 20 min and the cells were treated as described above. The first batch of viral particles were harvested 48 h post-transfection and fresh media was added to the cells. A second batch of viral particles were harvested 72 h post-transfection. Both batches, were pre-cleared by centrifugation at 200 *g* for 10 min at room temperature and filtered through a 0.45-µm filter before immediate use in transduction.

2.4.2 Transduction of HEK-293T and CPT cells with retroviral vectors

HEK-293T cells and CPT cells were seeded 18 h prior to infection with MVV single cycle viruses. Unless stated otherwise, $0.5x10^5$ cells were seeded in a 48-well plate in a total volume of 200 µl per well and were infected with virus preps, typically, with the virus quantity corresponded to 0.5 mU of associated RT activity. 2 d post-infection, cells were expanded into 6-well plates, and infected cells were passaged 1:2 -1:10 5 d post-infection. Cells were harvested 7 d post-infection.

For infections with MLV-derived vectors, $7x10^6$ LHKO cells were seeded in T175 flasks (Corning) 18 h prior to transduction. Freshly-prepared viral suspensions were added to the cells; 2 d post second round of transduction, growth medium was supplemented with 1 µg/ml puromycin for selection of stably transduced cells.

2.4.3 Transfection of CPT-3 cells

CPT-3 cells were transfected with gRNA encoding CRISPR/Cas9 plasmid, as previously described (Ran et al., 2013). $4x10^6$ CPT-3 cells were seeded in complete medium 16 to 18 h prior to transfection in 10 cm dishes. 15 µl Xtreme Gene 9 transfection reagent was first added to Opti-MEM, followed by 5 µg of pSpCas9(BB)-2A-GFP-gRNA1/2/3 in a total volume of 500 µl. The transfection mixture was incubated at room temperature for 15 min. The complete medium on the seeded cells was replaced with Opti-MEM and the transfection mixture was added drop-wise to the cells.

2.5 Biochemistry

2.5.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were separated on SDS-PAGE gels, composed of a 4% stacking gel layered over an 11% resolving gel. The resolving gel was prepared using 11% (w/v) acrylamide, 0.29% (w/v) bisacrylamide, 0.1% (w/v) SDS and 0.38 M Tris-HCl, pH 8.8, supplemented with 0.042% (w/v) ammonium persulphate (APS) and 0.84% (v/v) N,N,N',N'-tetramethylethane-1,2-diamine (TEMED). The stacking gel mix was prepared with 4% (w/v) acrylamide, 0.1% (w/v) bisacrylamide, 0.1% (w/v) SDS and 0.12 M Tris-HCl, pH 6.8, supplemented with 0.046% (w/v) APS and 0.123% (v/v) TEMED. In some instances, commercial 4 - 20% gradient Tris-glycine PAGE gels (Bio-Rad) were used. Proteins and Laemmli sample buffer (Table 2-1) were mixed in 1:1 ratio and incubated at 95ºC for 10 - 15 min. 10 - 12 µl sample was loaded per well and electrophoresis was conducted at 100 V in TGS buffer, prepared from 10x stock solution (Table 2-1).

2.5.2 Coomassie staining

SDS-PAGE separated proteins were visualised by staining with Instant*Blue* colloidal Coomassie Blue staining solution (Sigma-Aldrich). Stained gels were imaged using a Gel Doc EZ gel documentation system (Bio-Rad).

2.5.3 Western blotting

2.5.3.1 ***Preparation of whole cell extracts***

Cells, harvested by trypsinisation, were washed twice with PBS and lysed in freshly prepared RIPA buffer containing 150 mM NaCl, 1% Triton-X, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris pH 8.0 and two pellets (to obtain 2x recommended final concentration) of cOmplete EDTA-free protease inhibitor cocktail (Sigma-Aldrich) for 20 min at 4ºC. The lysates were pre-cleared by centrifugation at 21,000 *g* for 30 min at 4ºC to remove any insoluble material. The total amount of protein in the lysates was quantified using bicinchoninic acid assay (BCA, Pierce). The samples were diluted in Laemmli buffer to equal total protein amounts prior to separation by SDS-PAGE. Typically, 10 - 20 µg of total protein per lane was used for Western blotting.

2.5.3.2 ***Electrotransfer and immunodetection***

Proteins were separated by SDS-PAGE were transferred on to Trans-Blot Turbo Mini Polyvinylidene difluoride (PVDF) membranes (Bio-Rad) using a Transblot Turbo semi-dry electro-transfer unit (Bio-Rad) for 30 min at 25 V. To minimize non-specific antibody binding during the next steps of the protocol, the membranes were blocked overnight at 4ºC in 5% (w/v) skimmed milk powder (Marvel) in PBS supplemented with 0.1% Tween-20 (PBS-T). The membranes were then incubated with respective primary antibodies, diluted in blocking solution, for 1.5 h at room temperature with gentle shaking. The membranes were washed with three changes of PBS-T for a total of 30 - 45 min to remove unbound primary antibodies. The membranes were then incubated with horseradish peroxidase (HRP)-conjugated or a fluorophore-tagged secondary antibodies in blocking solution for 1 h with gentle shaking and washed with three changes of PBS-T.

Depending on the secondary antibody used, the protein detection on the membranes was carried out appropriately, using either chemiluminescence or fluorescence. For detection of chemiluminescence, the membranes were incubated in ECL-prime reagent (GE Healthcare), with equal parts of solution A and solution B (both supplied with the kit), as advised by the manufacturer. The membranes were then visualised

using Imager 600 RGB (Amersham). Fluorescence was detected directly on the membranes using Odyssey infrared light detection system (LI-COR).

2.5.3.3 ***Stripping membranes for Western blotting with alternative antibodies***

For re-probing with alternative antibodies, membranes were extensively washed with PBS-T prior to incubation in ReBlot Plus Milk Antibody Stripping Solution (prepared from a 10x concentrate, Sigma-Aldrich) for 15 min at room temperature. Stripped membranes were washed with PBS-T and incubated overnight at 4ºC in blocking solution prior to re-use in Western blotting.

2.5.3.4 ***Antibodies used for Western blotting***

The various antibodies used in this study are listed in Table 2-5.

Table 2-5 Details of the antibodies used in this study

No.	Name	Description	Dilution
Primary Ab			
1	αLEDGF/p75	Polyclonal rabbit anti-LEDGF/p75 antibody, Bethyl Laboratories (A300-848A)	1 in 5,000
2	αHRP2	Polyclonal rabbit anti-HRP2 antibody, Novus Bio (NBP2-47438)	1 in 1,000
3	αActin	Monoclonal HRP conjugated anti-Actin (13E5) antibody, Cell Signaling Technology (5125)	1 in 1,000
4	αCA (p24)	Rabbit anti-MVV CA (p24) antibody, Engelman lab	1 in 5,000
Secondary Ab			
1	αRabbit	Polyclonal Goat anti-Rabbit IR Dye 800 CW Ab, Li-Cor (925-32211)	1 in 2,000
2	αRabbit HRP	Polyclonal Goat anti-Rabbit HRP conjugated Ab, BioRad (1706515)	1 in 10,000

2.5.4 Bacterial expression and purification of WT and mutant MVV IN proteins

WT and mutant forms of MVV IN were recombinantly expressed using the modified pCPH6P bacterial expression vector in a His_6- tagged form (Figure 2-1) (Cherepanov, 2007). *E. coli* PC2 cells were transformed with pCPH6P-MVV IN (WT or appropriate mutant), seeded on LB agar Petri dishes supplemented with 120 µg/ml ampicillin and incubated at 30ºC overnight. A single colony was used to inoculate 100 - 300 ml LB broth supplemented with 120 µg/ml ampicillin, and bacteria were allowed to grow overnight with vigorous shaking at 30ºC.

Three to twelve flasks, each containing 1 L LB broth with 120 µg/ml ampicillin were inoculated with 25 ml overnight culture and incubated at 30ºC for 2 - 3 h in a shaking

incubator at 250 rpm. The optical density at 600 nm (OD_{600}) of the culture was measured, and when the OD_{600} value reached a value between 0.6 and 0.9, protein expression was induced by addition of 0.01% (w/v) IPTG (Table 2-1). The bacterial cultures were incubated at 25ºC for 6 h at 250 rpm, and the cultures were pelleted by centrifugation at 6,000 rpm for 10 min at 4ºC. The bacterial pellets were stored at -80ºC until further use.

For purification, bacterial pellets were thawed and resuspended in core buffer, which contained 1 M NaCl, 25 mM BisTris-propane (BTP)-HCl, pH 7.0 (Table 2-1) and 7.5 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), supplemented with 0.2 mM phenylmethylsulfonyl fluoride (PMSF, Table 2-1). CHAPS is a zwitterionic detergent, which increases solubility of MVV IN (Cherepanov, 2007, Ballandras-Colas et al., 2017). Bacteria were disrupted by sonication using Sonifier 450 (Branson) on ice at power level 7 for 2 min at a time, allowing the cells to cool down in between several (4-6) consecutive sonication rounds. The lysate was clarified by centrifugation at 18,000 rpm (JA-25.50 rotor, Beckman Coulter) for 1 h at 4ºC. The supernatant was supplemented with 15 mM imidazole (Table 2-1) and incubated with 3 ml pre-washed Ni-NTA resin (Qiagen) for 30 min at 4ºC by gentle rocking. The supernatant and beads were then applied to a disposable Econo-Pack column (Bio-Rad) and supernatant allowed to flow through the beads by gravity, the beads were washed with four changes of 25 ml core buffer supplemented with 15 mM imidazole to remove unbound proteins. His_6-tagged MVV IN was eluted with 10 ml core buffer containing 200 mM imidazole in 1 ml fractions and immediately supplemented with 5 mM DTT. The fractions were analysed by SDS-PAGE gel and protein bands were visualised by staining with colloidal Coomassie Blue. Fractions containing recombinant protein of interest were pooled. The total protein content was estimated by Bradford method, using bovine serum albumin (BSA) standard solution (supplied as a 2 mg/ml stock, Pierce). To cleave His_6 affinity tag, the protein was incubated with human rhinovirus 14 (HRV14) 3C protease (laboratory stock; 1 mg protease per 50 mg of tagged protein) was added and the protein was incubated overnight at 4ºC.

The following day, digested protein was analysed by SDS-PAGE and staining with colloidal Coomassie Blue to verify removal of the affinity tag. Recombinant protein

was diluted 1:8 in salt free buffer (core buffer without salt) to adjust NaCl concentration to 120 mM NaCl and injected into a 5-ml HiTrap Heparin column (GE Healthcare). The protein was eluted by a linear 0.1 - 1 M NaCl gradient in core buffer, using an ÄKTA purifier system (GE Healthcare), monitoring A_{280}, A_{260} and conductivity. 1-ml fractions were collected and supplemented with 5 mM DTT and 370 mM NaCl, to adjust salt concentration to ~1 M. Fractions were analysed on SDS-PAGE and staining with colloidal Coomassie Blue. Fractions containing recombinant protein were pooled and concentrated to 10 - 30 mg/ml, using a VivaSpin 20 centrifugal concentrator with a 10-kDa cut-off (VivaProducts). Concentrated MVV IN was aliquoted appropriately (typically 25-50 µl per aliquot), flash frozen in liquid nitrogen and stored at -80ºC until further use.

2.5.5 Bacterial expression and purification of LEDGF/p75

LEDGF/p75 was expressed using the modified pCPH6P bacterial expression vector in a His_6-tagged form (Cherepanov, 2007). *E. coli* PC2 cells were transformed with pCPH6P-LEDGF/p75 plasmid, and colonies were grown on LB agar Petri dishes containing with 120 µg/ml ampicillin at 30ºC overnight. A single colony was used to inoculate 300 ml LB broth supplemented with 120 µg/ml ampicillin, and bacteria were grown overnight with vigorous shaking at 30ºC.

Twelve flasks, each containing 1 L LB broth with 120 µg/ml ampicillin were inoculated with 25 ml overnight culture and shaken at 30ºC for 2 - 3 h in a shaking incubator at 250 rpm. When the OD_{600} value reached a value between 0.6 and 0.9, protein expression was induced by addition of 0.01% (w/v) IPTG. The cultures were grown and pelleted as described in section 2.5.4. The bacterial pellets were stored at -80ºC until further use.

For purification, bacterial pellets were thawed and resuspended in core buffer, which contained 1 M NaCl, 50 mM Tris-HCl pH 7.5, supplemented with 0.2 mM PMSF. The cells were lysed and clarified as mentioned in section 2.5.4. The supernatant was supplemented with 10 mM imidazole and incubated with 3 ml Ni-NTA resin (Qiagen) for 30 min at 4ºC with gentle rocking. The supernatant and beads were then applied to an disposable Econo-Pack column (Bio-Rad) and supernatant allowed drain by gravity, the beads were washed with four changes of 25 ml ice-cold core buffer

supplemented with 5 mM imidazole to remove unbound proteins. His_6-tagged LEDGF/p75 was eluted with 10 ml core buffer containing 200 mM imidazole in 2-ml fractions. Fractions were supplemented with 10 mM DTT, and analysed by SDS-PAGE and staining with colloidal Coomassie Blue. Fractions containing LEDGF/p75 were pooled, the total protein content was estimated by Bradford method using BSA standard solution (Pierce) and treated with HRV14 3C protease overnight at 4ºC, as described in section 2.5.4.

The following day, the protein was analysed by SDS-PAGE gel and staining with colloidal Coomasie Blue. The recombinant protein was diluted 1:8 in low-salt buffer (Buffer A, 25 mM Tris-HCl, pH 7.5) to adjust salt concentration to 120 mM and injected on to a 5-ml HiTrap SP column (GE Healthcare). The protein was eluted by in a linear 0.1 - 1 M NaCl gradient, while monitoring A_{280}, A_{260}, and conductivity in 1.5-ml fractions. Each 1.5-ml fraction was supplemented with 2 mM DTT and analysed by SDS-PAGE and colloidal Coomassie Blue staining.

LEDGF/p75-containing fractions were pooled and concentrated to a volume of 5 ml, using a VivaSpin-20 centrifugal concentrator with 30-kDa cut-off (VivaProducts). The concentrated protein was injected into a HiLoad Superdex 200 column (GE Healthcare), pre-equilibrated with gel filtration buffer (GFB, 25 mM Tris, pH 7.5, 0.5 M NaCl). Size exclusion chromatography was conducted at a flow rate of 1-2 ml/min; 2-ml fractions were collected and supplemented with 2 mM DTT. Fractions containing full-length LEDGF/p75, free of degradation products, were pooled, concentrated and aliquoted. The purified protein was flash frozen in liquid nitrogen and stored at -80ºC until further use.

2.5.6 Multi Angle Laser Light Scattering (MALLS)

MVV IN proteins were diluted to a final concentration of 8, 4, 2, and 1 mg/ml in IN dilution buffer (1 M NaCl, 7.5 mM CHAPS, 2 mM DTT, 25 mM BTP-HCl, pH 7.0). 45 µl diluted IN was filtered through a 0.22-µm Amicon Ultra spin filter (Sigma-Aldrich) by centrifugation at 4ºC for 30 sec at 10,000 *g* to remove any aggregated material. The full volume of each filtered protein sample was transferred into a 250-µl sample vial (Thermo Fisher, catalogue #C4011-13), and stored in a refrigerated cabinet at 4ºC prior to automatic injection. 30 µl protein was injected on to a Superdex 200

Increase 3.2/300 column (GE Healthcare), equilibrated in 1 M NaCl, 3 mM NaN_3, 25 mM BisTris-HCl, pH 6.5. Size exclusion chromatography proceeded at a buffer flow rate of 50 µl/min, while monitoring A_{280}, refraction and light scattering using in-line UV flow cell, differential refractometer and multi-angle laser light scattering photometer. The latter directed a laser at the solution to detect intensity of the scattered light at 16 angles. The total scattering intensity is related to the molecular weight of the particles in solution, with higher protein oligomers displaying higher scattering intensity. The average molecular weight of the protein particles was determined from the total scattering of the protein solution and the protein concentration, estimated from the refraction index. The light scattering data were analysed using the ASTRA software version 6.1 (Wyatt Technology Corp., Santa Barbara, CA) (Caswell, 2019).

2.5.7 *In vitro* strand transfer assay

MVV IN proteins and LEDGF/p75 were adjusted to appropriate concentrations in IN diluting buffer (B_{IN}, 1 M NaCl, 10 mM BTP, pH 7.0, and 2 mM DTT) and LEDGF/p75 diluting buffer (B_L, 0.5 M NaCl, 10 mM BTP, pH 7.0, and 2 mM DTT), respectively. The strand transfer reactions were carried out in a total volume of 40 µl. A master mix was prepared with 40 mM KCl, 5 mM $MgSO_4$, 5 µM $ZnCl_2$, 2.5 mM DTT, 7.5 ng/µl pGEM plasmid, and 25 mM BisTris-HCl, pH 6.0. 34 µl of master mix was supplemented with 2 µl of 20 µM vDNA or water (for vDNA-omit control), 2 µl 0.5 - 1 mg/ml IN (or B_{IN} for IN-omit control), and 2 µl 1 - 2 mg/ml LEDGF/p75 (or B_L for LEDGF/p75-omit control) to obtain a total reaction volume of 40 µl. vDNA was prepared by annealing oligonucleotides EV272 and EV273 in 200 mM NaCl, 50 mM Tris-HCl, pH7.4.

The reactions were incubated at 37ºC for 1 h and terminated by addition of 4 µl stop solution containing 250 mM EDTA and 5% SDS. Reaction were supplemented with 30 µg (1.5 µl at 20 mg/ml) proteinase K (Thermo Scientific) and incubated further for 1 h at 37ºC to allow complete digestion of IN and LEDGF/p75. DNA components, including reaction products, were precipitated by the addition of 2.5 volumes (114 µl) 100% ethanol. Following incubation on ice for 15 min, the samples were centrifuged at 21,000 *g* for 30 min at 4ºC and supernatants was discarded. DNA pellets were air-

dried before the addition of 20 µl 1.2x DNA loading dye and analysed by electrophoresis through 1.5 % agarose gels in TAE buffer.

2.5.8 Quantification of FS integration strand transfer products by qPCR

DNA from strand transfer reactions was diluted 1:2,500 in Milli-Q water. 1 µl diluted DNA was combined with PC1235 and PC1237 PCR primers (each at the final concentration of 0.25 µM) and PowerUp SYBR-Green master mix (Thermo Fischer) in a total reaction volume of 20 µl. A standard curve was generated using an upscaled WT IN strand transfer reaction, serially diluted from 1:2,500 to ~1:10^{-7} in Milli-Q water. Each qPCR reaction was carried out in triplicate. Reactions were assembled in MicroAmp Optical 96-well reaction plates and sealed with MicroAmp Optical Adhesive film (Applied Biosystems).

The PCR was carried out in a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems) at 95ºC for 20 sec (hold stage), followed by 40 cycles of 95ºC for 30 sec, 68ºC for 1 min (PCR stage). The relative quantities of strand transfer products were calculated based on a standard curve generated using QuantStudio 7systems software (Applied Biosystems).

2.5.9 *In vitro* intasome assembly

The standard intasome assembly mixture comprised 40 mM potassium acetate, 3 mM $CaCl_2$, 1 mM DTT, 10 µM $ZnCl_2$ and 3.75 µM vDNA (prepared by annealing EV272 or EV272-5′-Cy3 and EV273; the former oligonucleotide was 5′-Cy3-modified to obtain labelled vDNA), 100 µg LEDGF/p75 (10 µl 10 mg/ml stock, prepared in B_L) and 45 µg MVV IN (9 µl 5 mg/ml stock prepared in B_{IN}), 25 mM BisTris-HCl, pH 6.0, in a final volume of 200 µl. Following incubation at 37ºC for 13 min, the assembly mixture was supplemented with an additional 175 mM NaCl (to adjust the final salt concentration to 310 mM) and 50 mM BisTris-HCl, pH 6.5.

The nucleoprotein complexes were separated from unassembled protein and vDNA by size exclusion chromatography through a Superdex 200 Increase 10/300 GL column (GE Healthcare) in 310 mM NaCl, 3 mM $CaCl_2$, and 50 mM BisTris-HCl, pH 6.5.

2.5.10 Subcellular fractionation

Subcellular fractionation or chromatin extraction was carried out to detect the presence of residual LEDGF/p75 in CPT-3 cells depleted of LEDGF/p75, as previously described (Llano et al., 2006). $10x10^6$ cells were harvested by trypsinisation and twice washed with PBS. The cells were lysed in 1 ml cold cytoskeletal I buffer (CSK I), composed of 10 mM Pipes, pH 6.8, 100 mM NaCl, 1 mM EDTA, 300 mM sucrose, 1 mM $MgCl_2$, 1 mM DTT, 0.5% Triton X-100, 1 mM PMSF and two pellets (to obtain 2x recommended final concentration) of cOmplete EDTA-free protease inhibitor cocktail (Sigma-Aldrich) for 15 min at 4ºC. $1/10^{th}$ of the lysate (total fraction, T) was mixed with RIPA buffer, while the remained cell lysate was centrifuged at 500 *g* for 3 min at 4ºC. The supernatant (S1 fraction) contained triton-soluble proteins. The pellet was resuspended in CSK II buffer, composed of 10 mM Pipes, pH 6.8, 50 mM NaCl, 300 mM sucrose, 6 mM $MgCl_2$, 1 mM DTT, 0.5% Triton X-100 and two pellets (to obtain 2x recommended final concentration) of cOmplete EDTA-free protease inhibitor cocktail. The samples were treated with DNase I for 30 min, followed by extraction with 250 mM NH_2SO_4 for 10 min at 25ºC. The DNase- and salt-treated sample was then centrifuged at 1,200 *g* for 6 min at 4ºC. The resultant supernatant (S2 fraction) contained DNase-released chromatin-associated proteins. The process henceforth was carried out as described in section 2.5.3.

2.6 Virology

2.6.1 Virus quantification using Taq-PERT assay

Viruses were quantified by a TaqMan probe-based product enhanced reverse transcriptase (Taq-PERT) qPCR assay (Li et al., 2015), based a procedure originally reported by Vermeire and colleagues (Vermeire et al., 2012).

5 µl viral stock was combined with 5 µl lysis buffer (2 µl RNase inhibitor added to 98 µl lysis buffer immediately prior to use, Table 2-1) and incubated at room temperature for 10 min. The lysed virus was diluted with 90 µl core buffer (Table 2-1). The qPCR reaction comprised of 12.5 µl TaqMan Gene Expression Master Mix (supplied as a 2x concentrate; Thermo Fisher), primers PC1335 and PC1336 (each at the final

concentration of 0.5 µM), TaqMan probe PC1566 (at a final concentration of 0.2 µM), 0.125 µl RiboLock RNase Inhibitor (Thermo Fisher), 0.125 µl phage MS2 phage RNA template (Sigma-Aldrich), 2 µl diluted and lysed virus and 9.95 µl Milli-Q water to adjust the total reaction volume to 25 µl. The standard curve was generated using recombinant HIV-1 RT (Merck Millipore; supplied at a concentration of 10,000 mU/µl). Prior to use, HIV-1 RT was serially diluted in core buffer to 4, 0.8, 0.16, 0.032, and 0.0064 mU/µl. 2 µl each diluted sample was added to the PCR reaction. A mock reaction containing core buffer and no RT or virus was used as a negative control for the qPCR. Reactions were assembled in MicroAmp Optical 96-well reaction plates and sealed with MicroAmp Optical Adhesive film (Applied Biosystems).

The PCR was carried out in QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems, Thermo Fisher) at 42ºC for 20 min (RT step), 50ºC for 2 min, 95ºC for 10 min (Hot Start), 95ºC for 15 sec, 60ºC for 1 min, cycled 40 times. All reactions were performed in triplicate. The relative viral quantities were calculated based on the standard curve generated using QuantStudio 7 systems software (Applied Biosystems).

2.6.2 Fluorescence activated cell sorting (FACS)

Flow cytometry was used to count cells infected with a GFP reporter virus or transfected with a GFP reporter plasmid. Adherent cells, harvested by trypsinization, were washed with PBS and treated in 6.5% (w/v) formaldehyde in PBS. Fixed cells were centrifuged at 200 *g* for 10 min at room temperature and resuspended in 200 - 500 µl PBS. 10,000 single cells were analysed for GFP expression using a Fortessa flow cytometer (BD Biosciences). GFP was excited at 495 nm and emission was detected with a 530/30 band pass filter. The data was acquired using BD Diva software (BD Biosciences) and analysed in FlowJo v13. Live cells (population P1) were initially gated using the area plot of forward scatter (FSC-A) versus side scatter (SSC-A), to separate them from cell debris. Singe cells (population P2) were then differentiated from doublets in population P1 by gating on FSC-A vs forward scatter height (FSC-H). 10,000 single cells were analysed for GFP expression by gating for 530/30 blue and FSC-H.

Single-cell sorting on live transfected CPT-3 was done as follows. Cells were trypsinized, washed with PBS and resuspended in FACS buffer (1% FBS, 1 mM EDTA in PBS, sterilized by filtering through a 0.22-µm membrane) at $4x10^6$ cells/ml in polypropylene tubes. The cells were transferred on ice to the Flow Cytometry facility (The Francis Crick Institute) and aseptically sorted using MoFlow XDP instrument with a 100-µm nozzle. Single GFP-positive CPT-3 cells were sorted into 96-well plates, containing 120 µl media (1:1 ratio of fresh media and pre-conditioned media, harvested from confluent culture of CPT-3 cells and filtered through a 0.22-µm pore membrane).

2.6.3 Luciferase assay

To measure expression of firefly luciferase, cells were harvested by trypsinization and washed twice with PBS and centrifuged at 200 *g* for 10 min at room temperature. The cell pellet was resuspended in 200 µl PLB (prepared from a 5x concentrate; buffer composition is given in Table 2-1). It was placed in dry ice for 10 min followed by incubation in 37ºC for 10 min. The lysed sample was centrifuged at maximum speed in a table-top centrifuge (21,000 *g*) for 30 min at 4ºC and the clear supernatant was stored at -80ºC until further use. 20 µl thawed cellular extract was combined with 100 µl luciferase assay reagent (LAR, Table 2-1) and loaded into a well of an OptiPlate-96 (Perkin Elmer). Mock samples containing only LAR and PLB were also loaded to measure background luciferase readings. The readings were normalized to the total protein concentrations in the samples, which was measured using bicinchoninic acid (BCA) method. To this end, 1 ml of BCA reagent (prepared by mixing 50 parts of solution A and 1 part of solution B, supplied with BCA protein assay kit; Pierce) was combined with 10 µl cellular extract or PLB (mock) and the samples were incubated at 42ºC for 30 min. 200 µl of each sample was loaded into flat-bottom, transparent 96-well plates and light absorption was measured at a wavelength of 562 nm (A_{562}) using SpectraMax plate reader (GMI). The blank measurement was subtracted from the absorbance measured of the samples. The luciferase measurements were normalised to the total protein content, thereby generating Relative Luciferase Units (RLU = Luciferase reading/A_{562}).

2.6.4 Quantification of vDNA synthesis by PCR amplification of the *luc* gene

18 h before infection, $0.35x10^6$ HEK-293T cells were seeded in 24-well plates in 400 µl complete medium per well. Prior to infection, virus (0.3 mU RT units) was treated with 0.12 U/µl DNase I (Invitrogen), for 1 h at 37ºC to minimize carryover of plasmid DNA. DNase-treated virus was added to the cells, and cells were harvested 2, 8 and 24 h post-infection. At an appropriate time point, cells were harvested by trypsinization, washed with PBS, centrifuged at 200 *g* for 30 min at 4ºC and stored at -80ºC as pellets. In control conditions, cells were pre-incubated with 150 µM AZT (100 mM AZT stock in DMSO, stored at -20ºC) for 1 h prior to infection with DNase-treated virus. 2 h post-infection, after the cells were washed with PBS, the 8- and 24-h cell samples were supplemented with fresh 100 µM AZT.

Cellular DNA was purified using Quick-DNA Microprep Kit (Zymo Research). For this purpose, cell pellets were lysed by addition of 400 µl Genomic Lysis buffer (supplied with the kit), vortexed for 6 sec and incubated at room temperature for 10 min. Lysed mixtures were transferred to Zymo-Spin IC columns with a collection tube (supplied with the kit) and centrifuged at 10,000 *g* for 1 min at room temperature. The columns were transferred to a fresh collection tube (supplied) and washed with 200 µl DNA Pre-Wash buffer (supplied). DNA retained on column filters was washed again with 500 µl genomic DNA Wash buffer (supplied) and spin columns were drained by centrifugation at 10,000 *g* for 1 min at room temperature. Finally, columns were transferred to fresh sample collection tubes and DNA was eluted with 15 µl of Elution Buffer (supplied) and stored at -20ºC for further use.

TaqMan probe-based qPCR was carried out with the DNA purified from infected cells. A master mix was prepared with PCR primers PC1488 and PC1489 (each primer at a final concertation of 0.1 µM), Taq-Man probe PC1490 (0.2 µM), TaqMan Gene Expression Master Mix (Thermo Fisher) and Milli-Q water. To obtain a total reaction volume of 20 µl, 16 µl of master mix was combined with 4 µl of DNA (adjusted to 25 ng/µl with Milli-Q water). A standard curve was generated with serial 10-fold dilutions of pCVW-CG-Luc plasmid, starting at a plasmid copy number of $5x10^7$ to $5x10^{-1}$. A mock reaction containing no DNA (Milli-Q water only) was used as a negative control for the qPCR. All qPCR reactions were carried out in triplicate.

qPCR was carried out in QuantStudio 7 Flex Real-Time PCR System at 50ºC for 2 min, 95ºC for 10 min, 95ºC for 15 sec, 60ºC for 1 min, 72ºC for 30 sec, cycled 40 times. The relative quantities of *luc* DNA in the samples were calculated based on the standard curve using QuantStudio 7 systems software (Applied Biosystems).

2.6.5 Quantification of viral DNA synthesis by PCR amplification of the late reverse transcription (LRT) product

LRT product was measured as an alternative method for quantification of vDNA synthesis. The method used TaqMan probe-based qPCR, using PCR primers amplifying a DNA fragment spanning 5′-LTR and the beginning of the *gag* gene. The LRT qPCR TaqMan probe was designed to anneal within the *gag* portion.

A standard curve was generated with serial 10-fold dilutions of pCR4-TOPO-MVV-LRT plasmid, starting at a plasmid copy number of $5x10^5$ to 5 copies/µl. A mock reaction with no DNA (Milli-Q water only) was used as a negative control for the qPCR. The reaction was carried out in QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems, Thermo Fisher) at 50ºC for 2 min, 95ºC for 10 min, 95ºC for 15 sec, 60ºC for 1 min, 72ºC for 15 sec, cycled 40 times. The relative vDNA quantities in the samples were calculated based on the standard curve generated using QuantStudio 7 systems software (Applied Biosystems).

The cells were infected with MVV vectors as described in section 2.4.2. Briefly, cells seeded in a 48-well plate, were infected with DNase-treated viruses (using 2.5 mU RT virus per well). HEK-293T cells were infected with WT virus in the presence of 100 µM ddC, similar to infections carried out in the presence of AZT (section 2.6.4). 8 h post-infection, infected cells were harvested and DNA was purified, using Quick-DNA Microprep Kit (Zymo Research), as described in section 2.6.4.

2 d post-infection, infected cells were processed as described in section 2.4.2. 7 d post-infection, DNA was purified from infected cells in 6 well plates, following the manufacturer's protocol, using GenElute Mammalian Genomic DNA Miniprep Kit (Merck). Cells, harvested by trypsinization, were centrifuged for 5 min at 300 *g* and the supernatant was discarded. Cells were resuspended in 200 µl Resuspension Solution (supplied) and treated with 20 µl RNase A solution (supplied) for 2 min at room temperature. The cells were lysed by adding 400 µg Proteinsase K solution

(supplied as 20 mg/ml) and 200 µl Lysis Solution C (supplied). The cell suspension was vortexed for 15 sec and incubated at 70ºC for 10 min. 200 µl ethanol (95-100%) was added to the lysate and vortexed to generate a homogenous mixture. The lysate was loaded into pre-assembled GenElute Miniprep Binding Column (supplied), pre-washed with 500 µl Column Preparation Solution (supplied), and centrifuged at 7,500 *g* for 1 min. The flow-through was discarded and the column was washed twice with 500 µl Wash Solution (supplied). The DNA was eluted with 200 µl Elution Solution (supplied) into a fresh tube and stored at -20ºC until further use. A TaqMan probe-based qPCR reaction was carried out with the DNA purified from infected cells, as described in section 2.6.4, with a standard curve generated using pCR4-TOPO-MVV-LRT plasmid.

2.6.6 Infections in the presence of INSTIs

$0.5x10^5$ HEK-293T, LKO and LHKO cells were seeded 18 h prior to infection in a 48-well plate in a total volume of 200 µl per well. DTG, dissolved in 100% DMSO at 10 mM, was diluted with Milli-Q water to 10 µM (0.1% DMSO). The drug was further serially diluted with complete medium to final concentrations of 10,000, 1,000, 100, 10, and 1 nM. Immediately before infection, the media in the plates was replaced with DTG-containing media (or complete media with 0.1% DMSO and no drug), and cells were infected with 0.5 mU RT units of WT-KV1772 or E154Q MVV vector. Infected cells were treated as described in section 2.4.2 and 7 d post-infection, luciferase activity was measured in cell extracts as described in section 2.6.3.

2.6.7 Optimisation of transfection of sheep choroid cells using nucleofection

Optimisation of nucleofection was performed using Cell Line Optimization 4D-Nucleofector X Kit (Lonza) following the Amaxa 4D-Nucleofectior optimisation protocol (Lonza). CPT-3 cells were trypsinized and washed with PBS. Three 1.5-ml tubes with $6x10^6$ CPT-3 cells each were centrifuged at 90 *g* for 10 min at room temperature. The cell pellets were resuspended in 400 µl SE, SF or SG 4D-Nucleofector X solutions and 8 µg pmaxGFP (8 µl, at a concentration of 1 µg/µl) was added per tube and mixed gently. 20 µl cells were added per well to a 16-well Nucleocuvette strip (supplied) and Experimental Set up with 15 test protocols was

performed in Amaxa 4D-Nucleofectior (Lonza). The Nucleocuvette strip was incubated at room temperature for 10 min. Next, the cells were resuspended in 80 µl pre-warmed complete medium and added to 48 well plate containing 150 µl per well. 2 d post-nucleofection, cells were processed for flow cytometry, as described in section 2.6.2. Based on the flow cytometry analysis of transfected cells, SE buffer was selected as the optimal Nucleofector solution for CPT-3 cells, along with EN-138 protocol for nucleofection.

2.6.8 Assembly and introduction of Cas9 RNPs into CPT-3 cells for *PSIP1* and *HRP2* gene knockouts

Alt-R Cas9 Electroporation Enhancer (IDT), crispr RNAs (crRNAs) targeting LEDGF/p75 or HRP2 and Alt-R CRISPR-Cas9 tracrRNA (IDT) were resuspended in IDTE buffer (10 mM Tris, pH 8.0 and 0.1 mM EDTA) to 100 µM. Guide RNAs (gRNAs) were prepared by heating equal quantities of crRNA and tracrRNA (a final concentration of 50 µM) at 95°C for 5 min and cooled to room temperature. Ribonucleoproteins (RNPs) were prepared by incubating 30 µM gRNA (6 µl of 50 µM stock) and 24.8 µM Alt-R S.p. Cas9 Nuclease V3 (4 µl of 62 µM stock, IDT) for 20 min at room temperature. The prepared gRNAs and RNPs were stored at -20°C.

CPT-3 cells were treated with trypsin and washed with PBS. $1.5x10^6$ cells were centrifuged at 90 *g* for 10 min at room temperature. The cell pellet was resuspended in 70 µl SE Cell Line Nucleofector Solution (Lonza) and 18 to 36 µl prepared RNP complexes were added along with 4 µM Alt-R Cas9 Electroporation Enhancer (4 µl of 100 µM stock, IDT) in a total reaction volume of 110 µl, adjusted with PBS. 100 µl of the total suspension was added to the Nucleofection chamber (Lonza) and placed in the Amaxa 4D-Nucleofectior (Lonza). The chambers were treated with the nucleofector program EN-138 and incubated at room temperature for 10 min. The nucleofected cells were gently resuspended in 300 µl complete medium and added to 6-well plate with 2 ml complete medium per well. 2-3 d post-nucleofection, nucleofected cells were expanded 1:10 into 10 cm dishes and the remainder were treated as described in section 2.5.3 and analysed for LEDGF/p75 and HRP2 expression by Western blotting.

2.6.9 Thin-section EM

HEK-293T cells were transfected with plasmids to generate MVV WT-KV1772 viral vectors in twelve 15-cm dishes and harvested by ultracentrifugation, as described in section 2.4.1. The protocol for visualisation of the viral particles was previously described (Wight et al., 2014). The viral pellet was resuspended in sterile PBS and incubated at 4ºC for 20 min. Equal volume of primary fixative (5% glutaraldehyde and 200 mM sodium cacodylate) was added to the viral suspension (to achieve a final concentration of 2.5% glutaraldehyde and 100 mM sodium cacodylate) and incubated on ice for 60 min. The fixed viral suspension was centrifuged at 17,800 *g* for 60 min at 15ºC. The viral pellet was resuspended in 2% low melting temperature agarose (Sigma-Aldrich) and incubated at 37ºC for 15 min followed by centrifugation at 16,200 *g* for 20 min at 38ºC. The sample was incubated on ice for ~3 h and topped up with 2.5% glutaraldehyde and 100 mM sodium cacodylate and stored at 4ºC overnight for the agarose to harden.

The hardened agarose was scooped out of the tube and washed in milli-Q water. Excess agarose around the pellet was cut using a single edge razor blade (VWR). The agarose with the viral pellet was placed in a 7 ml glass vial (Agar Scientific). The following steps were carried out in a chemical fume hood at the Electron Microscopy Facility (The Francis Crick Institute). The Milli-Q water was exchanged with 1% (w/v) osmium tetroxide (using sufficient volume to cover the complete agar slice) and incubated for 90 min at room temperature. The osmium tetroxide was exchanged with milli-Q water and washed 2x in Milli-Q water. The Milli-Q water was exchanged with 1% (w/v) aqueous uranyl acetate and incubated by gently rocking for 90 min at room temperature. The uranyl acetate was exchanged with Milli-Q water and washed 2x with Milli-Q water. The agarose with the viral pellet was gradually dehydrated in an ethanol series, by first incubating in 50% ethanol for 10 min, followed by incubation with 70% ethanol for 10 min, 90% ethanol for 10 min and finally 3x in 100% ethanol for 10 min each. The ethanol dehydration was followed by 2x 100% propylene oxide dehydration steps for 10 min each.

Epon 812 resin (TAAB Laboratories) was prepared following manufacturer's instructions. Briefly, 10.15 g of Epon 812, 2.5 g of dodecenylsuccinic anhydride (DDSA), 7.6 g of methyl-5-norbornene-2,3-dicarboxylic anhydride (MNA) and 0.4 g

of tris (dimethylaminomethyl)phenol (DMP-30) were mixed gently on a magnetic stirrer for 1-3 h, without generating air bubbles. The samples were treated with 2 parts propylene oxide and 1 part resin for 20 min, followed by equal parts of propylene oxide and resin for 30 min and 1 part propylene oxide with 2 parts resin for 30 min and finally in 100% resin for 10 min. The samples were then stored at 4ºC overnight. Thin slices of the agar containing the viral pellet were either flat embedded on Aclar film slides (Agar Scientific) or in flat embedding moulds (Agar Scientific) and topped up with fresh preparation of the Epon 812 resin. The samples were left to polymerise at 70ºC overnight. From this point, the samples were processed by Dr. Raffaella Carzaniga (Electron Microscopy Facility, The Francis Crick Institute).

The resin blocks were sectioned serially using a UC7 ultramicrotome (Leica Microsystems), from which 70-nm sections were picked on to formvar-coated slot copper grids (Gilder Grids). The sections were counterstained with 2% (w/v) aqueous uranyl acetate solution, followed by Reynold's lead citrate for 5 min (Reynolds, 1963). The sections were then viewed using a Tecnai G2 Spirit Transmission Electron Microscope operating at 120 kV tension (Thermo Fischer Scientific) and images were acquired using Oneview CCD camera (Gatan/Ametek). The images analysed using Fiji open source platform (Schindelin et al., 2012).

2.6.10 Genomic DNA for mapping MVV integration sites in human and sheep cells

MVV WT viruses were produced in twelve 15-cm dishes each, as described in section 2.4.1. 18 h before infection, $0.35x10^6$ HEK-293T, CPT-3, LKO and LHKO cells were seeded in 6-well plates. Fresh viral preparations were used to infect the cells and 2 d post-infection, the cells were expanded into 10-cm dishes. 5 d post-infection, a small fraction of cells were fixed and GFP expression was measured using flow cytometry (section 2.6.2). The remainder of the cells were harvested to purify genomic DNA, using GenElute Mammalian Genomic DNA Miniprep Kit (Merck), as described in section 2.6.5. Isolated DNA was processed further for integration site sequencing by Dr. Parmit Singh (Engelman laboratory, Dana-Farber Cancer Institute, Boston, USA) (Serrao et al., 2016). The sequencing data were processed

by Dr. Parmit Singh, Engelman lab and Prof. Peter Cherepanov (described in Serrao et al., 2016), and the genomic integration site maps were generated.

Chapter 3. Probing the functional significance of the expanded intasome architecture in Maedi-visna virus

3.1 Aims

The cryo-EM structure of the homo-hexadecameric MVV intasome was previously determined in our group (Ballandras-Colas et al., 2017). As discussed in section 1.5.4, the largest intasome assembly observed thus far, it harbours unique structural features that have been observed only in lentiviral IN proteins. These include, a CTD bridge forming intra- and inter-tetramer interactions and the α-helical CCD-CTD linker.

The aim for this part of the project is to test the importance of the hexadecameric MVV intasome assembly by targeting its unique features *in vitro*. Amino acid substituted MVV IN proteins were expressed, purified and their biochemical properties were studied by a range of techniques. The enzymatic functions of the mutants were measured in an *in vitro* IN strand transfer assay and compared with WT MVV IN, along with assessing their ability to assemble into stable intasomes.

3.2 Design rationale for amino acid substitutions in MVV IN

A tetrameric intasome assembly is sufficient to catalyse the integration reaction, as seen in PFV (spumavirus), HTLV-I and STLV (deltaretroviruses) (Hare et al., 2010, Maertens et al., 2010, Barski et al., 2020, Bhatt et al., 2020). In order to test the importance of the hexadecameric intasome assembly of MVV IN, amino acid substitutions were introduced to disrupt the CTD-CTD interface interactions in the CTD bridge and destabilise the α-helical CCD-CTD linker. Crucially, the features targeted by mutagenesis were specific to the lentiviral intasome, and the mutations were designed to minimize disturbing the interfaces involved in the CIC assembly (described in section 1.5.5, Figure 1-12). It is important to note that due to the homo-hexadecameric nature of the IN complex within the MVV intasome, all sixteen IN subunits possessed the introduced substitutions.

At selected positions, as indicated in figure 3-1A, residues interacting through their protruding hydrophobic side chains were targeted for substitutions. Residues participating in the intra-tetramer interactions, Trp-245, Val-252, Tyr-261 and Val-263 were substituted with either Ala or Glu. In case of Trp-245 and Val-252, substitutions were also made with Leu and Asp, respectively. Similarly, residues participating in inter-tetramer interactions, Phe-223, Tyr-225 and Ile-272 were substituted with either Ala or Glu. It was hoped that introducing Ala and Leu substitutions would reduce the extent of hydrophobic interactions within these regions, while substitutions with the acidic residues (Asp and Glu) would lead to a like-charge repelling effect (Zhou and Pang, 2018).

The second unique feature of the lentiviral intasome assembly, α-helical CCD-CTD linker, was also targeted. Although the CCD-CTD linker of MVV IN is of an intermediate size (~20 amino acid residues) compared with that of PFV IN (60 amino acid residues) and MMTV IN (~10 amino acid residues), it assumes a compact α-helical configuration. This necessitates the requirement of two additional tetramers to contribute their CTDs to complete the CIC (described in section 1.5.5). Interestingly, although there is little to no sequence conservation within this region, all lentiviral CCD-CTD linkers are predicted to be α-helical (Ballandras-Colas et al., 2017). To test the importance of this α-helix, mutations were introduced in the MVV IN CCD-CTD linker to perturb its native conformation. Thus, stretches of Gln residues at positions 207 and 208, and from positions 211 to 213 were substituted with Pro and Gly residues to generate IN mutants QQ207GP and QQQ211PGG (Figure 3-1B). Unlike Gln, Pro and Gly disfavour formation of α-helical secondary structure (Hider and Hodges, 1984). Additionally, insertions were created with the PFV IN CCD-CTD linker sequence, residues 304 to 314, known to assume an extended conformation within the PFV intasome (Hare et al., 2010), from MVV IN position 211 to 213 (the mutant designated QQQ211PFVlinker).

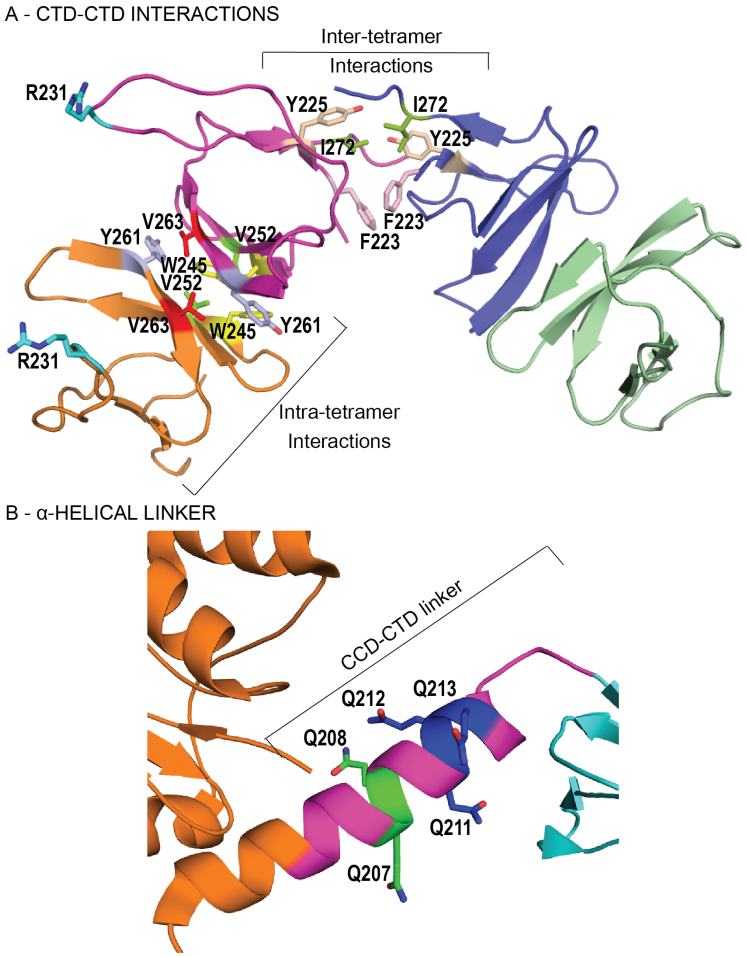

Figure 3-1 Structural features of the MVV intasome targeted by site-directed mutagenesis in this work.

(A) The CTD bridge, depicting intra- (orange & magenta) and inter- tetramer interactions (magenta and blue). Protein backbone is shown as cartoon and selected amino acid residues as sticks: F223 (pink), V263 (red), Y261 (purple), I272 (green), Y225 (beige), V252 (green) and W245 (yellow); R231 (cyan) as a CTD control. (B) The CCD-CTD α-helical linker (magenta). The targeted residues are indicated by sticks, QQ207 (green) and QQQ211 (blue).

Three control mutants were also designed. Introducing substitutions in any one of the carboxylates comprising the invariant D,D-35-E motif completely abrogates 3′-processing and strand transfer activities of IN (Engelman and Craigie, 1992, Kulkosky et al., 1992). In HIV-1, IN E152A was shown to efficiently synthesise full-length vDNA during reverse transcription, suggesting this is defective specifically for integration (Shin et al., 1994). In MVV, Glu-154 is structurally and functionally equivalent to HIV-1 IN Glu-152 (Hare et al., 2009a, Ballandras-Colas et al., 2017). Based on these observations, MVV IN E154Q was designed as an enzyme active site substitution. Zn^{+2} binding is coordinated by the invariant HHCC motif in the NTD, which is essential for IN multimerisation. H12N mutation in HIV-1 IN disrupted tetramerisation and the mutant protein behaved as a monomer – dimer mixture (Hare et al., 2009a). Thereby, based on the published observations in HIV-1 IN, a tetramerisation defective MVV IN H12N was designed as a second control. Additionally, Arg-231 (R231E) was targeted; this residue is located within the CTD, but outside of any of the CTD-CTD interfaces. Within the MVV intasome, Arg-231 residues from different IN subunits of the hexadecamer participate in interactions with vDNA and target DNA (Ballandras-Colas et al., 2017). The nomenclature of the amino acid substitutions was based on their positions in full-length MVV WT IN (residues 1-300). The amino acid substitutions, their positions and the locations of the targeted residues are listed in Table 3-1.

Table 3-1 Structure based amino acid substitutions of MVV IN

No.	Name	Description	Interaction
1	E154Q	Substitution in the CCD	DDE (active site)
2	H12N	Subsitution in the NTD	HHCC motif
3	R231E	Substitution in the CTD	CTD control
4	F223A	Substitution in the CTD	Inter-tetramer
5	V263E	Substitution in the CTD	Intra-tetramer
6	Y261A	Substitution in the CTD	Intra-tetramer
7	Y261E	Substitution in the CTD	Intra-tetramer
8	I272E	Substitution in the CTD	Inter-tetramer
9	Y225A	Substitution in the CTD	Inter-tetramer
10	V252A	Substitution in the CTD	Intra-tetramer
11	V252D	Substitution in the CTD	Intra-tetramer
12	W245E	Substitution in the CTD	Intra-tetramer
13	W245L	Substitution in the CTD	Intra-tetramer
14	W245A	Substitution in the CTD	Intra-tetramer
15	QQQ211PGG	Triple Gln substitution at position 211-213	CCD-CTD linker
16	QQ207GP	Double Gln substitution at position 207-208	CCD-CTD linker
17	QQQ211PFVlinker	Triple Gln substitution at position 211-213 with 11 amino acids from the long flexible PFV CCD-CTD linker	CCD-CTD linker

3.3 Expression and purification of recombinant MVV INs and LEDGF/p75

Mutations in the IN region were introduced by site directed mutagenesis, described in section 2.2.2. IN proteins, WT and mutants, and LEDGF/p75 were expressed in *E. coli* using a modified pCPH6P vector encoding an HRV 3C protease site following the N-terminal hexahistidine (His_6) tag. The constructs were based on the KV1772 MVV isolate. Recombinant proteins were purified by immobilized metal affinity chromatography using nickel affinity resin and treated with HRV14 3C protease to remove the affinity tags. They were purified further by ion exchange (Figures 3-2 and 3-3). In case of LEDGF/p75, an additional step of size exclusion chromatography was included (Figure 3-4). Further details on the purification methods are in sections 2.5.4 and 2.5.5.

The majority of mutant MVV IN variants could be produced and purified in a similar fashion to WT IN, with their ion exchange chromatography profiles comparable with that of WT IN (Figure 3-3). Typically, 6 to 10 mg purified protein was obtained from 3 L bacterial culture. Purified proteins could be concentrated to 30 mg/ml in a high-salt protein storage buffer. However, two of the mutant INs displayed a tendency to form precipitated aggregates, thereby requiring larger expression cultures for their purification. Thus, despite best efforts, only 1.5 mg H12N IN was produced, and the protein could only be concentrated to ~15 mg/ml; while V252A IN could not be concentrated above 5 mg/ml, with the total yield of only 0.8 mg. Purified recombinant proteins were aliquoted, flash frozen and stored at -80ºC. Although MVV IN appears stable and displays enzymatic activity even after multiple freeze thaw cycles (unpublished observations), due to the unpredictable behaviour of the mutant INs, all proteins assayed in this work were thawed once, and the aliquots were discarded after their first use.

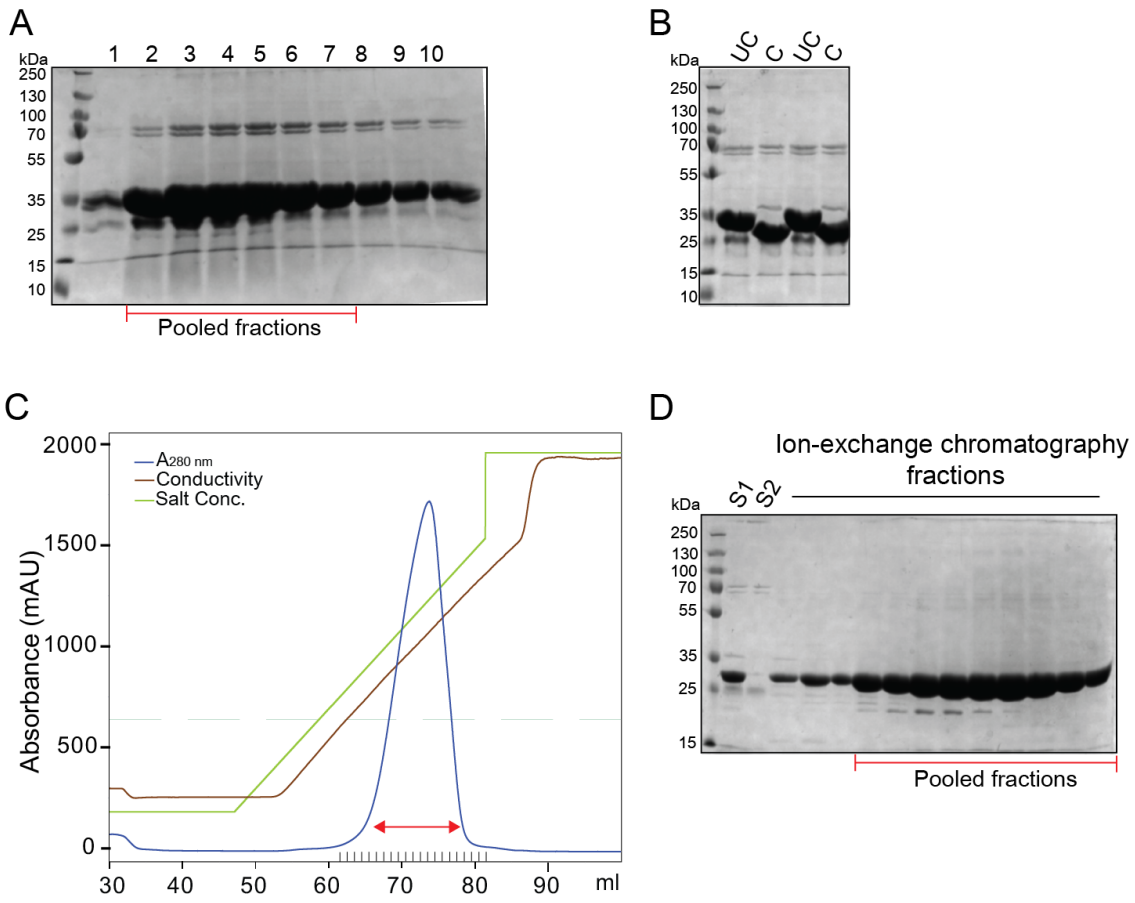

Figure 3-2 WT MVV IN purification.

(A) Analysis of proteins eluted from Ni-NTA Sepharose by SDS-PAGE and staining with Coomassie Blue. Molecular weight makers were separated in the left-most lane; migration positions (kDa) of maker bands are indicated to the left. (B) UC (Uncleaved) and C (Cleaved, with HRV14 3C protease) fractions from pooled Ni-NTA fractions analysed by SDS-PAGE. (C) Chromatogram of WT MVV IN purification by ion-exchange chromatography. The major peak eluting at ~70 ml contained MVV IN. The peak was collected in 1-ml fractions (fractions 19-30). (D) SDS-PAGE of chromatography fractions. S1 is diluted protein (1:8 in low salt buffer) and S2 is the flow-through from injecting the diluted protein into 5-ml HiTrapQ Heparin column, with the pooled fractions from ion-exchange chromatography indicated.

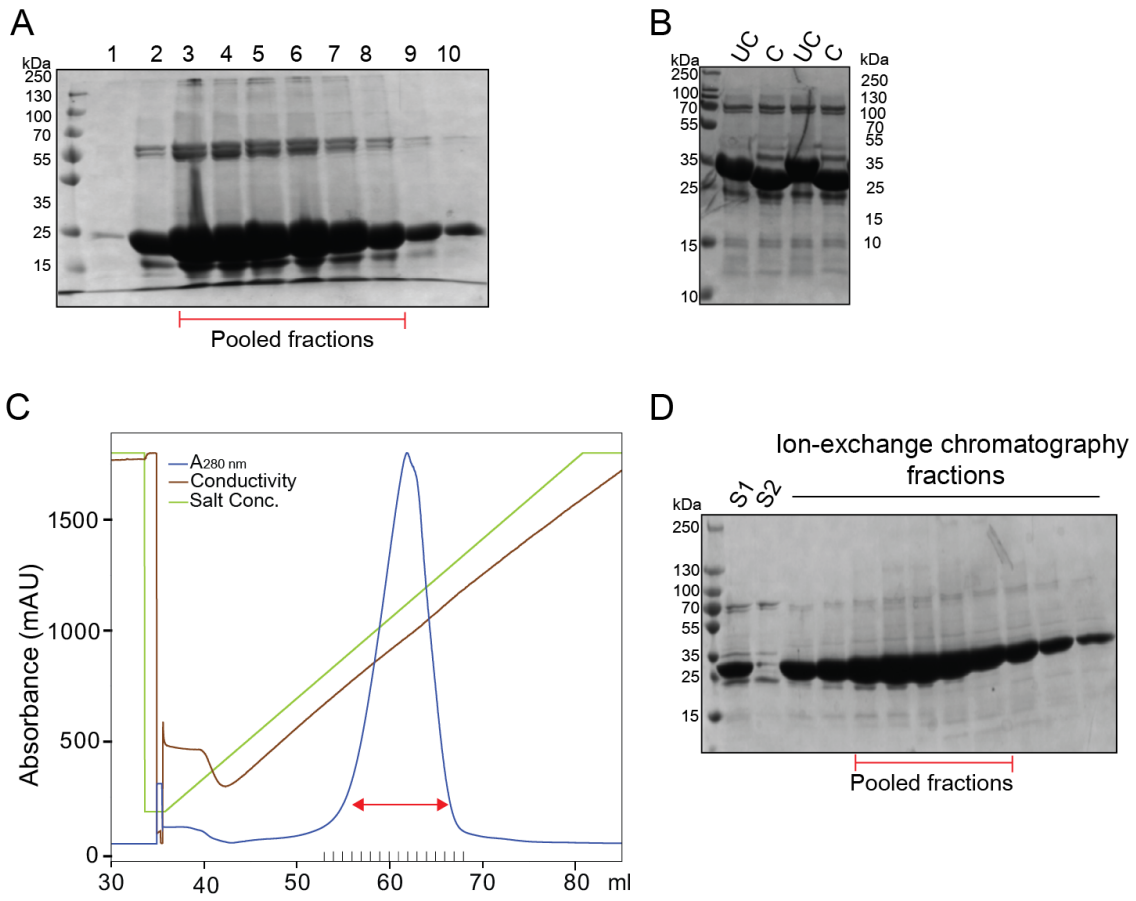

Figure 3-3 Y261A MVV IN purification.

(A) SDS-PAGE of proteins eluted from Ni-NTA beads. (B) SDS-PAGE of UC (Uncleaved) and C (Cleaved, with HRV14 3C protease) fractions from pooled Ni-NTA fractions. (C) Chromatogram from Y261A MVV IN purification by ion-exchange chromatography. The major peak eluting at ~60 ml contained MVV IN Y261A. The peak was collected in 1 ml fractions (fractions 14-23). (D) SDS-PAGE of chromatography fractions. S1 is diluted protein (1:8 in low salt buffer) and S2 is the flow-through from injecting the diluted protein into 5-ml HiTrapQ Heparin column, with the pooled fractions from ion-exchange chromatography indicated.

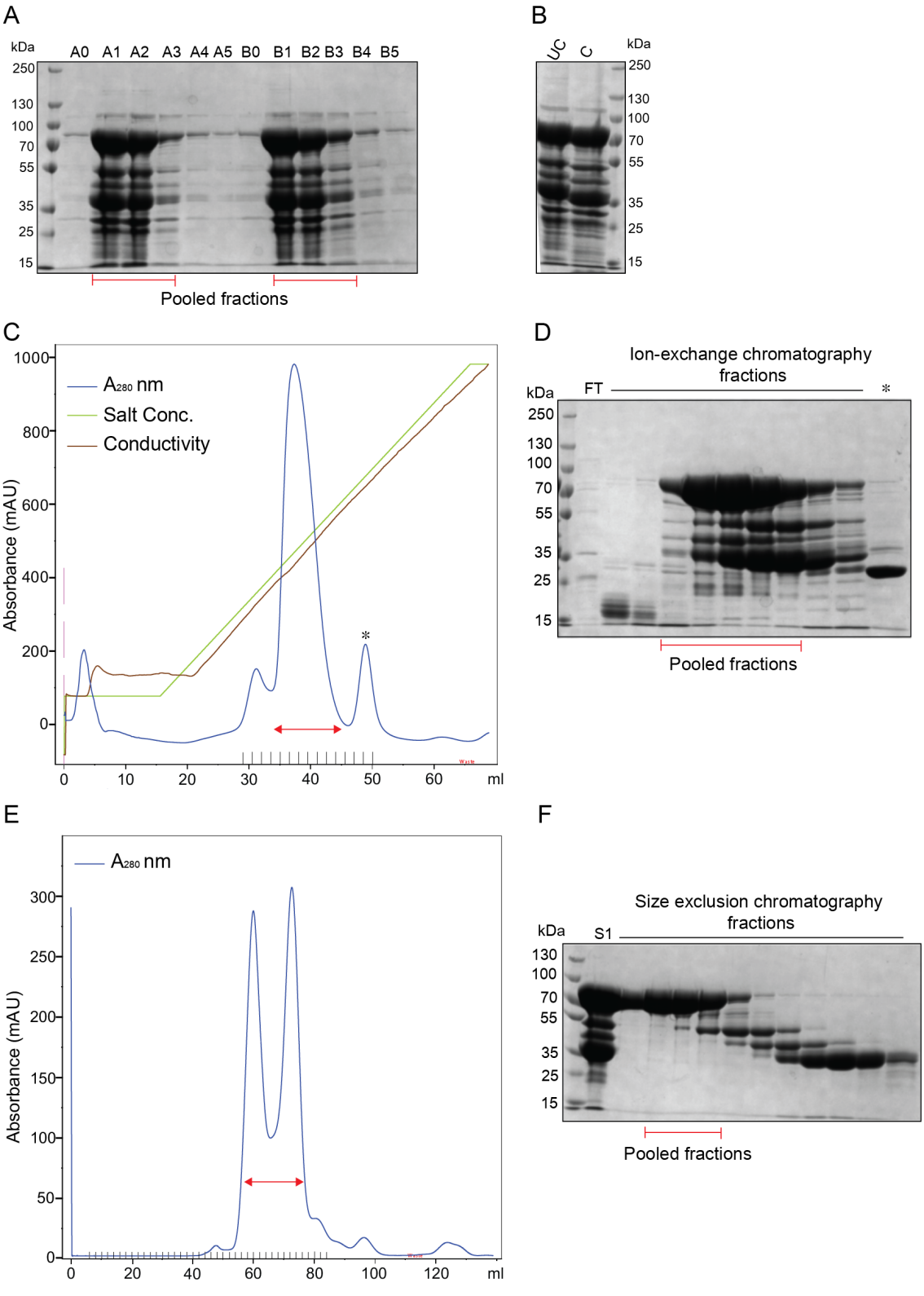
A
kDa
A0 A1 A2 A3 A4 A5 B0 B1 B2 B3 B4 B5
250
130
100
70
55
35
25
15
Pooled fractions
B
UC
C
kDa
250
130
100
70
55
35
25
15
C
1000
800
600
400
200
0
Absorbance (mAU)
A280 nm
Salt Conc.
Conductivity
*
0
10
20
30
40
50
60
ml
D
Ion-exchange chromatography fractions
kDa
FT
*
250
130
100
70
55
35
25
15
Pooled fractions
E
A280 nm
300
250
200
150
100
50
0
Absorbance (mAU)
0
20
40
60
80
100
120
ml
F
Size exclusion chromatography fractions
kDa
S1
130
100
70
55
35
25
15
Pooled fractions

Figure 3-4 Purification of recombinant LEDGF/p75.

(A) SDS-PAGE of proteins eluted from Ni-NTA beads after incubation with bacterial lysate. (B) SDS-PAGE of UC (Uncleaved) and C (Cleaved, with HRV14 3C protease) fractions from pooled Ni-NTA fractions. (C) Chromatogram from LEDGF/p75 purification by ion-exchange chromatography. The major peak eluting at ~40 ml contained LEDGF/p75, the minor peak (* eluting at ~50 ml) did not contain LEDGF/p75. The peak was collected in 2-ml fractions (fractions 15-23). (D) SDS-PAGE of chromatography fractions, FT is flow through from injecting the diluted protein into 5-ml HiTrapQ Heparin column, with the pooled fractions from ion-exchange chromatography indicated. (E) Chromatogram from LEDGF/p75 purification by size exclusion chromatography. LEDGF/p75 eluted with a retention volume of ~60 ml. (F) SDS-PAGE of chromatography fractions, S1 is the concentrated protein injected into the gel filtration column, with the pooled fractions from size exclusion chromatography indicated.

3.4 Multimerisation of MVV IN proteins in solution

Lentiviral INs in general and MVV IN in particular assemble into tetramers in solution and have a tendency to form higher order oligomers (Cherepanov et al., 2003, Faure et al., 2005, Hare et al., 2009a, Ballandras-Colas et al., 2017). Size exclusion chromatography with in-line multi-angle laser light scattering (SEC-MALLS) was used to determine the oligomeric states of WT and mutant MVV IN proteins. This technique allows to measure average molecular masses of species present in solution (Amartely et al., 2018).

The proteins were injected into a Superdex-200 HR column at 1, 2, 4, and 8 mg/ml. At 4 mg/ml, WT, R231E and E154Q displayed similar elution volumes of ~1.25 ml, while H12N eluted much later at ~1.6 ml. The behaviour of the control mutants was as expected. Indeed, R231E and E154Q do not alter any of the protein-protein interfaces within IN multimer, while H12N disrupts the NTD, and this domain is essential for higher-order multimer formation. All the remaining INs eluted between these two extremes (Figure 3-5A and B). A similar pattern was observed with all the tested concentrations of INs (Appendix Figures 8-1, 8-2 and 8-3). The experiments revealed that WT, E154Q and R231E INs formed the largest multimers, while H12N formed the smallest species in solution.

The minimum mass formed by WT MVV IN at the lowest input still allowing for detection of the protein elution using the differential refraction index detector (~1 mg/ml at the point of injection) was a tetramer (Figure 3-6), which was in agreement with published findings (Ballandras-Colas et al., 2017). At higher input concentrations, the protein formed larger species, with average apparent molecular mass approaching that of an octamer. As expected, the enzyme active site mutant E154Q IN as well as R231E IN behaved similar to WT protein. The remaining mutants formed intermediate-sized species ranging between a dimer and a tetramer at the lowest input concentration and displayed reduced abilities to form higher order multimers at higher input concentrations.

Measured molecular mass of H12N, the tetramerization defective mutant, unsurprisingly ranged between monomers and dimers, which is also in agreement with the reported observations made with HIV-1 IN H12N mutant (Hare et al., 2009a). Unfortunately, aggregation of H12N, V252A and QQQ211PFVlinker INs prevented MALLS analysis due to increased background in light scattering, especially at concentrations above 4 mg/ml. However, the elution behaviour of these proteins at the highest injected concentrations clearly indicated a defect in multimer formation (Figure 3-5).

The average molar masses of the inter-tetramer mutants F223A, Y225A and I272E ranged between 100 kDa and 130 kDa, with reduced abilities to oligomerise at the tested concentrations (Figure 3-6A). The intra-tetramer mutants, W245E, W245L, W245A, V252A, V252D, Y261E, Y261A and V263E displayed a range of multimerisation states in solution. Although two of the Trp mutants, W245E and W245A, did not form tetramers at 1 mg/ml, they assembled into higher order oligomers at the highest input concentration (8 mg/ml). The hydrophobic amino acids, Val-252 and Tyr-261 were substituted with either hydrophobic Ala (V252A and Y261A) or the acidic residues, Asp and Glu (V252D and Y261E). Interestingly, only the substitutions with the acidic residues caused severe tetramerisation defects (Figure 3-6B). This suggested the critical role of hydrophobic residues at these positions for tetramerisation of IN subunits in solution. MVV IN Val-263 is structurally equivalent to HIV-1 IN Val-260, substitutions of which were associated with multimerisation defects (Kalpana et al., 1999). This is in agreement with the observations made with MVV IN V263E (Figure 3-6B). The CCD-CTD linker mutants, QQ207GP, QQQ211PGG and QQQ211PFVlinker were also defective for tetramerisation, and only assembled into dimers (Figure 3-6A). Collectively, the results confirmed that mutations of MVV IN residues found within the CTD-CTD interfaces as well the α-helical CCD-CTD linker can strongly suppress self-association of the protein.

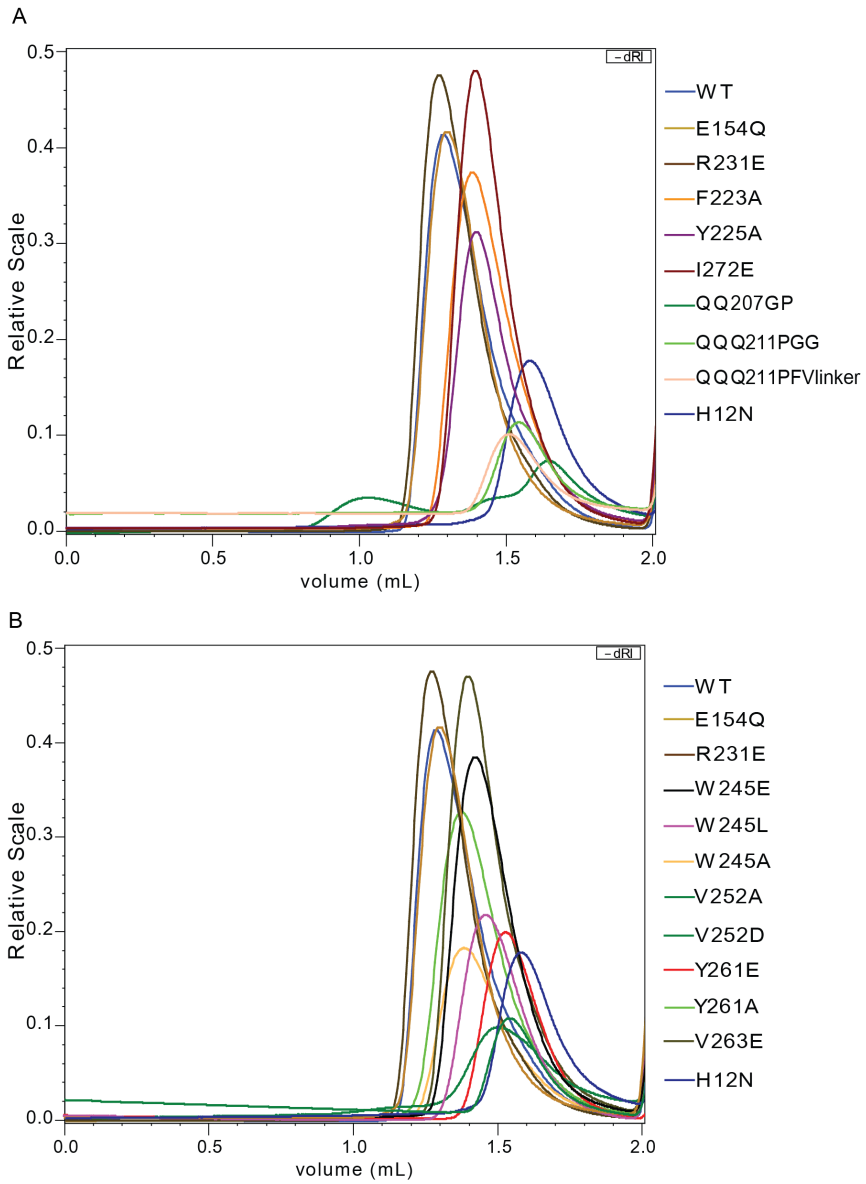

Figure 3-5 Size exclusion chromatography of MVV IN proteins.

Chromatograms from WT and mutant IN subunits injected at 4 mg/ml into Superdex 200 Increase 3.2/300 column, equilibrated in 1 M NaCl, 3 mM NaN3, 25 mM BisTris-HCl, pH 6.5. Elution profiles of IN WT, E154Q, H12N and R231E with (A) Inter-tetramer mutants – F223A, Y225A and I272E, CCD-CTD linker mutants – QQ207GP, QQQ211PGG and QQQ211PFVlinker, (B) Intra-tetramer mutants – W245E, W245L, W245A, V252A, V252D, Y261E, Y261A and V263E. Chromatograms of the proteins at 1, 2 and 8 mg/ml are in Appendix, Figures 8.1, 8.2 and 8.3, respectively.

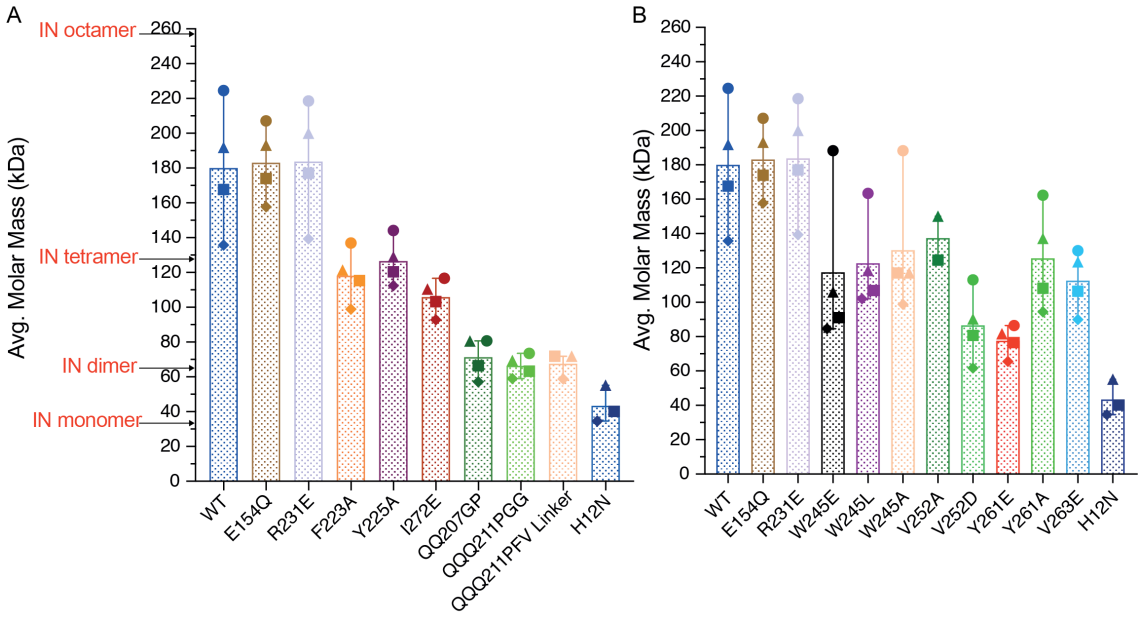

Figure 3-6 Oligomerisation properties of MVV IN variants determined by SEC-MALLS.

Comparison of the average molar masses of the IN subunits in solution, injected at 1, 2, 4, and 8 mg/ml and determined by SEC-MALLS. WT, E154Q, R231E and H12N MVV INs along with the variants carrying mutations within (A) inter-tetramer CTD-CTD interfaces: F223A, I272E and Y225A, and CCD-CTD linker mutants: QQ207GP, QQQ211PGG and QQQ211PFVlinker; (B) intra-tetramer CTD-CTD interfaces V263E, Y261E, V252A, V252D, W245E, W245L, W245A, and Y261A. The molecular masses of IN monomer (32 kDa), IN dimer (64 kDa), IN tetramer (128 kDa) and IN octamer (256 kDa) are indicated with arrow heads. Bar graphs plotted with means of the average molar masses of proteins in solution with the range. The concentrations of injection are indicated by symbols, diamond (1 mg/ml), square (2 mg/ml), triangle (4 mg/ml) and circle (8 mg/ml).

3.5 Strand transfer activities of the MVV IN proteins

Recombinant MVV IN, produced in *E. coli*, is capable of carrying out strand transfer activity *in vitro* in a LEDGF/p75-dependent manner (Ballandras-Colas et al., 2017). The strand transfer assays were performed in the presence of Mg^{2+} with purified recombinant MVV IN, LEDGF/p75 and synthetic double-stranded oligonucleotides mimicking pre-processed U5 vDNA ends. Supercoiled pGEM plasmid served as target DNA (Section 2.5.7). Strand transfer reactions were performed using varying IN and LEDGF/p75 inputs.

Integration of oligonucleotide substrates into the supercoiled plasmid target yields two major types of strand transfer products. Full-site (FS) product, also known as concerted integration product, is formed when a pair of oligonucleotide substrates are inserted into the target DNA. The formation of this product is thought to mimic the integration of a pair of vDNA ends during infection. Concerted integration of oligonucleotides into opposing strands of the target plasmid leads to its linearisation. Consequently, the FS product migrates similar to linearized plasmid during agarose gel electrophoresis. Multiple full-site strand transfer events lead to further fragmentation of the target DNA, producing downward smears on agarose gel. Conversely, single vDNA integration results in a branched half-site (HS) product (Figure 3-7A), which migrates close to the open circular form of the plasmid on the agarose gel.

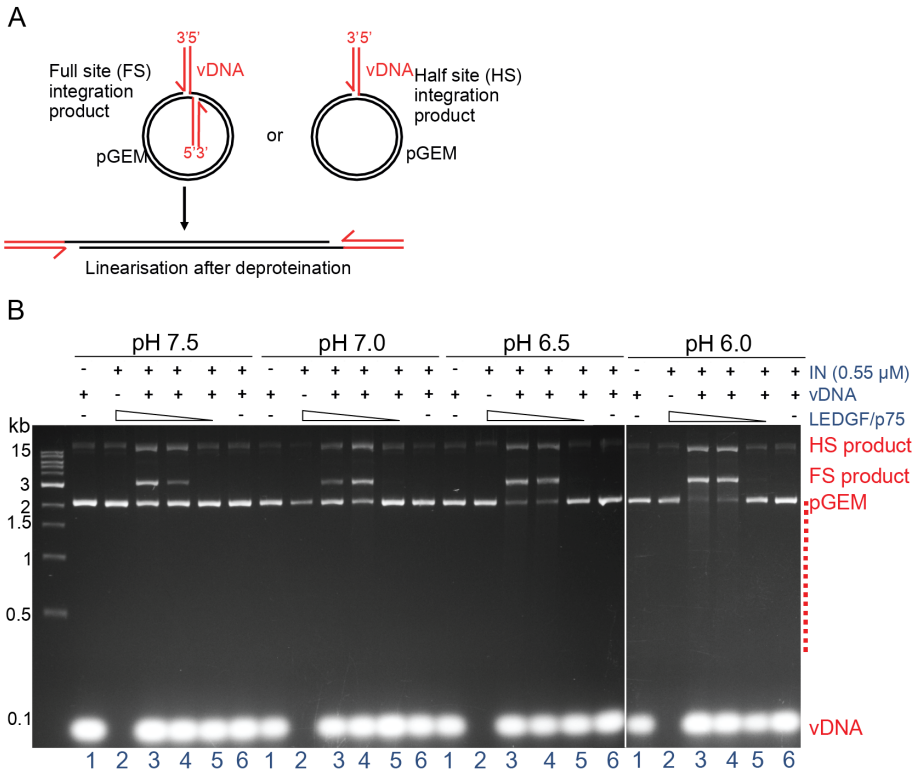

Figure 3-7 Optimisation of the MVV IN strand transfer assay.

(A) Schematic of *in vitro* integration reaction utilizing a plasmid DNA and double-stranded oligonucleotides as target and vDNA, respectively. Concerted insertion of two vDNA molecules into opposing strands of target DNA results in a linear full-site (FS) integration product (left). Half-site products arises from the insertion of one vDNA substrate molecule into one strand of the target plasmid (right). Red lines indicate vDNA and black lines are target DNA. (B) Optimal pH to carry out the strand transfer reactions was determined by incubating 0.55 μM WT IN with 1.5 μM (lanes 2 and 3), 0.75 μM (lane 4), 0.375 μM (lane 5) LEDGF/p75 or buffer (lane 6) with or without 20 μM vDNA, and 7.5 ng/ul pGEM (target DNA). Deproteinized reaction products were separated in 1.5% agarose gels and detected by staining with ethidium bromide. Migration positions of the reaction products, vDNA and pGEM are indicated on the right of the gel; positions of DNA molecular size markers (kb) are shown on the left. Lane 1 is a mock reaction that did not contain IN and LEDGF/p75, while the reaction in lane 2 did not contain vDNA. Lanes 3-5 contained all the reaction components, while LEDGF/p75 was omitted in lane 6.

3.5.1 Optimisation of *in vitro* strand transfer reaction

In agreement with published observations, WT MVV IN was highly active for strand transfer activity *in vitro*. A range of conditions were evaluated during optimisation of the strand transfer assay. In some instances, at excessive IN inputs, the target DNA was completely exhausted (Supplementary Figure 1, Ballandras-Colas et al., 2017) due to multiple full-site integration events. Several parameters, including IN and LEDGF/p75 input concentrations and pH of the reaction buffer, were trialled to determine the optimum conditions to carry out testing the IN strand transfer activity of the mutants. To determine the optimum pH to carry out *in vitro* IN strand transfer assays, reactions were carried out in parallel at four pH conditions, 7.5, 7.0, 6.5 and 6.0, using 0.55 μM WT IN with 1.5 or 0.75 or 0.375 μM LEDGF/p75 or buffer with 20 μM vDNA and 7.5 ng/μl pGEM target DNA (Figure 3-7B). Mock reactions lacking both proteins (IN and LEDGF/p75) or vDNA served as negative controls (Figure 3-7B, Lanes 1 and 2).

Under all the conditions tested, no FS integration products were detected in the absence of LEDGF/p75 nor in reactions with the lowest amount of LEDGF/p75 (0.375 μM) (Figure 3-7B, lanes 5 and 6). IN strand transfer activity increased with decreasing pH and, in agreement with published data (Ballandras-Colas et al., 2017), was highest at the lowest pH tested (pH 6.0), with products of multiple FS integration products present (Figure 3-7B).

Various concentrations of IN and LEDGF/p75 were used to carry out *in vitro* IN strand transfer reactions, using 2.2, 1.1, or 0.55 μM WT, E154Q and H12N IN. The reaction mixtures contained 3.0, 1.5, or 0.75 μM LEDGF/p75, or LEDGF/p75 dilution buffer in a mock control condition. Donor vDNA and pGEM target DNA were used at concentrations of 20 μM and 7.5 ng/μl, respectively. In agreement with published observations (Ballandras-Colas et al., 2017), WT IN displayed strand transfer activity *in vitro* only in the presence of LEDGF/p75 (Figure 3-8). IN and LEDGF/p75 inputs at 1.1 μM IN with 1.5 μM LEDGF/p75 and 0.55 μM IN with 0.75 μM LEDGF/p75 revealed the strongest FS product band, without completely exhausting the target DNA or a pronounced smearing due to multiple FS integration events (lanes 8 and 13, Figure 3-8). As expected, no integration products were observed with negative control IN mutants, E154Q and H12N under any of the conditions tested (Figure 3-8).

Furthermore, no target DNA degradation was observed in vDNA- or protein-omit conditions, verifying the absence of nuclease contamination in the reaction buffer components or recombinant proteins.

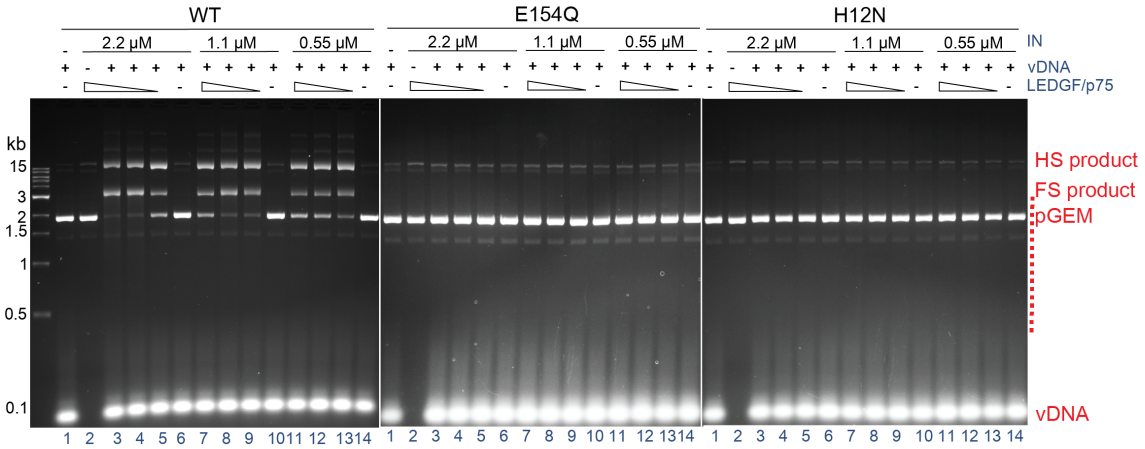

Figure 3-8 Strand transfer activities of WT, E154Q and H12N MVV IN proteins.

WT and mutant INs at a concentration of 2.2 μM (lanes 2-6), 1.1 μM (lanes 7-10), or 0.55 μM (lanes 11-14) were incubated with 3 μM (lanes 2, 3, 7 and 11), 1.5 μM (lanes 4, 8 and 12), or 0.75 μM (lanes 5, 9 and 13) LEDGF/p75 or buffer (lanes 6, 10 and 14), with or without 20 μM vDNA, in the presence of 7.5 ng/ul pGEM target DNA. Deproteinized reaction products were separated in 1.5% agarose gels and detected by staining with ethidium bromide. Migration positions of the reaction products, vDNA and pGEM are indicated on the right of the gel; positions of DNA molecular size markers (kb) are shown on the left. Lane 1 contained a mock reaction without IN and LEDGF/p75, lane 2 did not contain vDNA, lanes 3 and 4 contained all the components, and LEDGF/p75 was omitted in lane 5.

3.5.2 *In vitro* strand transfer activities of MVV IN mutants

The strand transfer activities of the mutant INs were tested *in vitro*, at concentrations of 1.1 or 0.55 μM. The reactions were carried out with 1.5 or 0.75 μM LEDGF/p75, 20 μM vDNA and 7.5 ng/μl target DNA. WT IN displayed robust strand transfer activity, manifesting as a smear of products of multiple FS integration events. As expected, active site mutant E154Q and tetramerisation defective mutant H12N were inactive. Remarkably, the majority of the novel IN mutants were highly compromised for the strand transfer activity. Five of the CTD mutants displayed greatly reduced, albeit still detectable levels of activity (Figure 3-9). Among these, Y261A was the most active mutant, followed by R231E, Y261E, F223A, and V263E, which was the least functional within the series. Of the three CCD-CTD linker mutants, QQQ211PGG was the only mutant with detectable levels of strand transfer activity.

All the mutations introduced to disrupt the CTD-CTD interface interactions and destabilise the CCD-CTD linker are outside the CIC, and thus are not expected to affect its assembly. However, the majority of these mutants were defective for strand transfer activity *in vitro*. In comparison to the WT IN, all of these mutants were unable to form tetramers and multimerise into higher order oligomers. These observations demonstrate the importance of multimerisation of INs into at least tetramers in solution in order to proficiently perform strand transfer *in vitro*.

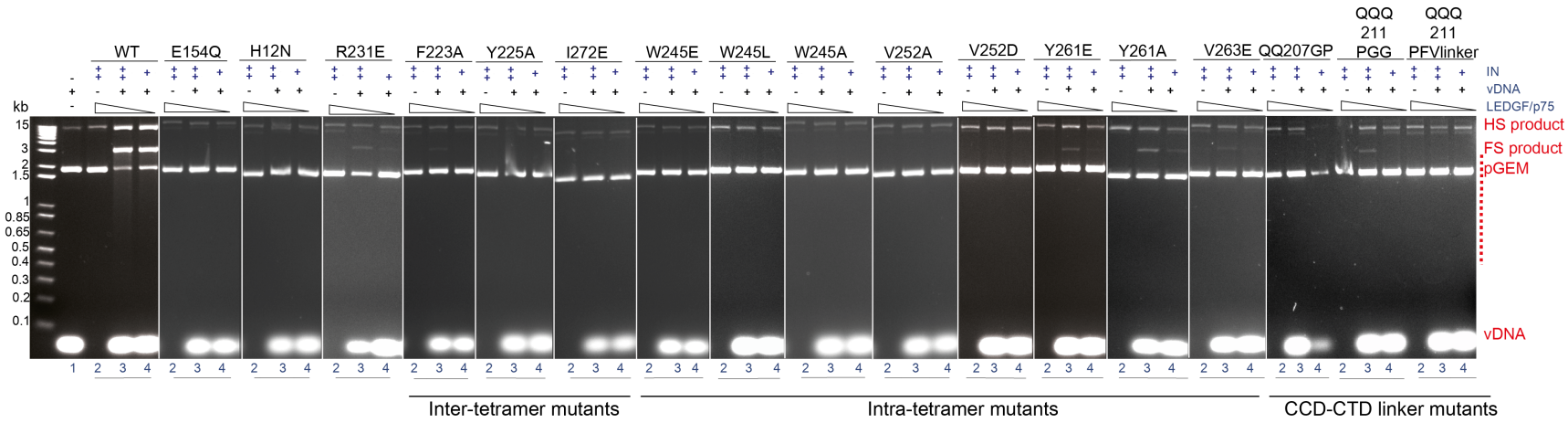

Figure 3-9 Strand transfer activities of MVV IN variants studied in this thesis.

WT and mutants INs at a concentration of 1.1 μM (lanes numbered 2 and 3) or 0.55 μM (lanes numbered 4) were incubated with 1.5 μM (lanes 2 and 3 ++) 0.75 μM (lanes 4, indicated by +) LEDGF/p75, with or without 20 μM vDNA, in the presence of 7.5 ng/ul pGEM target DNA. Deproteinized reaction products were separated in 1.5% agarose gels and detected by staining with ethidium bromide. Migration positions of the reaction products, vDNA and pGEM are indicated on the right of the gel; positions of DNA molecular size markers (kb) are shown on the left. Lanes 1 contained a mock reaction without IN and LEDGF/p75, lanes 2 did not contain vDNA, lanes 3 and 4 contained all the reaction components.

3.5.3 Quantification of strand transfer products

The ability of WT and mutant INs to perform strand transfer *in vitro* is detected on agarose gels. However, the presence of multiple FS integration events by WT and the faint FS integration product bands of the partially active IN mutants made it challenging to quantify the relative levels of strand transfer activity using images of ethidium bromide-stained agarose gels. To address this technical problem, the enzymatic IN activity was measured using real-time quantitative PCR (qPCR) with primers specific to the strand transfer products. The most preferable site for MVV integration into pGEM target plasmid, with a third of integration events occurring at this position, was previously established (A. Ballandras-Colas, unpublished observations). Therefore, PCR primers were designed to selectively amplify this vDNA - target DNA integrated junction, with one primer annealing within pGEM and the other complementary to the U5 vDNA end.

Strand transfer products from an upscaled reaction with WT MVV IN was diluted five-fold from 1:2,500 to 1:~10^7 in water and used in real time qPCR; DNA amplification was followed using SYBR green dye to generate a standard curve. This procedure was repeated with all subsequent strand transfer product quantification experiments.

The products from the reaction carried out in the presence of vDNA and target DNA and in the absence of both the proteins, IN and LEDGF/p75 (as indicated in lane 1, Figure 3-9) were used as negative control. The relative quantity seen for this mock reaction was considered a baseline for background amplification. Levels of strand transfer products obtained with each mutant were calculated as a proportion of those measured with WT IN, which was considered to be 100% (Figure 3-10). E154Q and H12N behaved similarly to the mock reaction. Y261A was the most active mutant with the highest quantity of amplified strand transfer product, corresponding to ~1.5% WT activity (Figure 3-10). Following this were R231E, V263E, F223A, and Y261E, which is the least active mutant. QQQ211PGG also displayed specific amplification for strand transfer products. Overall, results of quantification of the strand transfer activities by real-time PCR (Figure 3-10) agree well with conclusions based on comparing intensities of the FS products in ethidium bromide-stained gels (Figure 3-9).

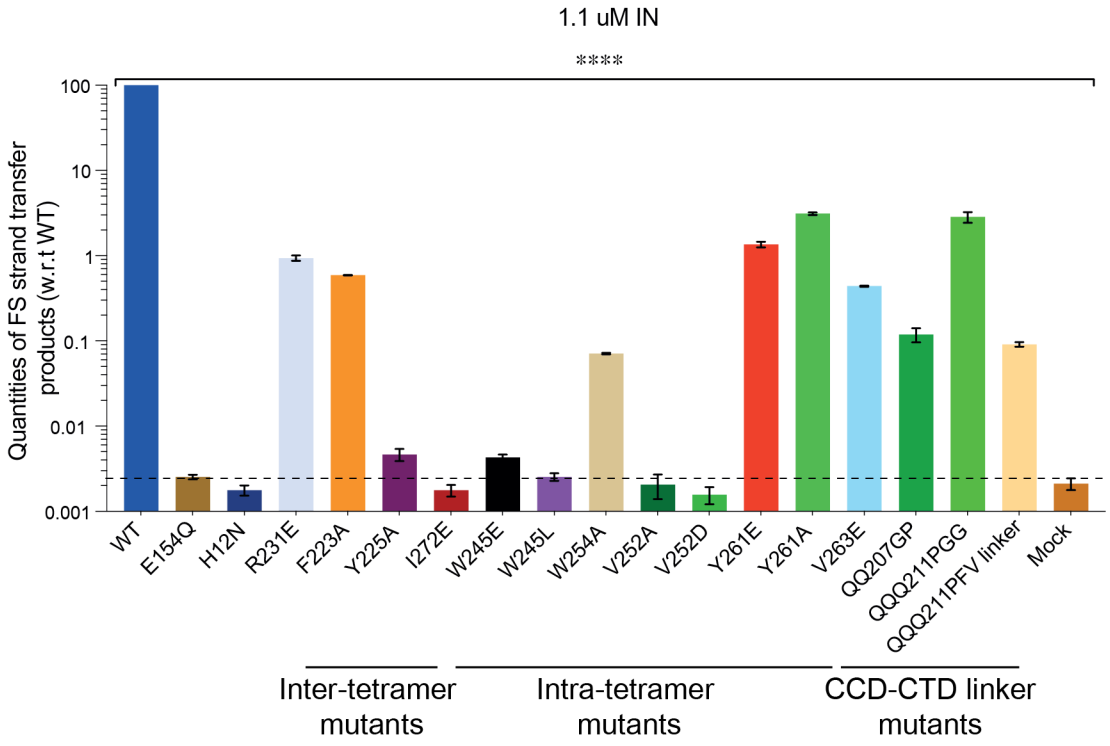

Figure 3-10 Relative strand transfer activities of MVV IN variants determined by quantitative real-time PCR.

Strand transfer reaction carried out with 1.1 μM IN (reaction products from lanes 3, Figure 3-9); Mean values plotted with SD from three qPCR reactions; the activity of WT MVV IN was set to 100. Statistical test was one-way Anova with multiple comparisons (**** = P <0.0001).

3.6 *In vitro* intasome assembly

IN must assemble into functional intasome to catalyse full-site integration (Wei et al., 1997, Hare et al., 2010). As observed in Figures 3-6 and 3-9, the IN proteins with mutations at the CTD-CTD interfaces and the α-helical CCD-CTD linker were greatly compromised for multimerization as well as strand transfer activity *in vitro*. A severe defect in the ability to form functional intasomes would be consistent with both of these observations.

In order to assess the ability of the INs to form intasome complexes, *in vitro* assembly reactions were carried out. WT and mutant MVV IN proteins were incubated with LEDGF/p75 and vDNA under conditions optimised for strand transfer activity, but omitting target DNA and Mg^{2+}. The absence of the essential cofactor Mg^{2+} prevented strand transfer reaction during assembly and chromatography. Only omitting plasmid DNA alone would not be sufficient, as the excess vDNA used in the assembly could serve as target DNA. The IN-vDNA complexes were separated from unassembled components (IN, LEDGF/p75 and free vDNA) by analytical size exclusion chromatography.

As expected, WT IN readily assembled into the intasome that eluted from the Superdex-200 column with an elution volume of 8.41 ml, followed by IN-LEDGF/p75 complex and vDNA leaving the size exclusion column at 8.83 ml and 15.22 ml, respectively (Figure 3-11A). Because the intasome contains both protein and DNA, it is characterized by near equal absorbance at wavelengths of 280 and 260 nm. As expected, MVV IN mutant E154Q likewise efficiently assembled into an intasome (Figure 3-11B). In addition, protein-DNA complex formation was observed with Y261A IN, which formed an intasome eluting at 8.44 ml (Figure 3-11C). Intasome assembly by the remaining IN mutants, could not be detected in these experiments, possibly due to poor separation of the nucleoprotein complexes from the excess protein complexes, as indicated by V263E (Figure 3-11D).

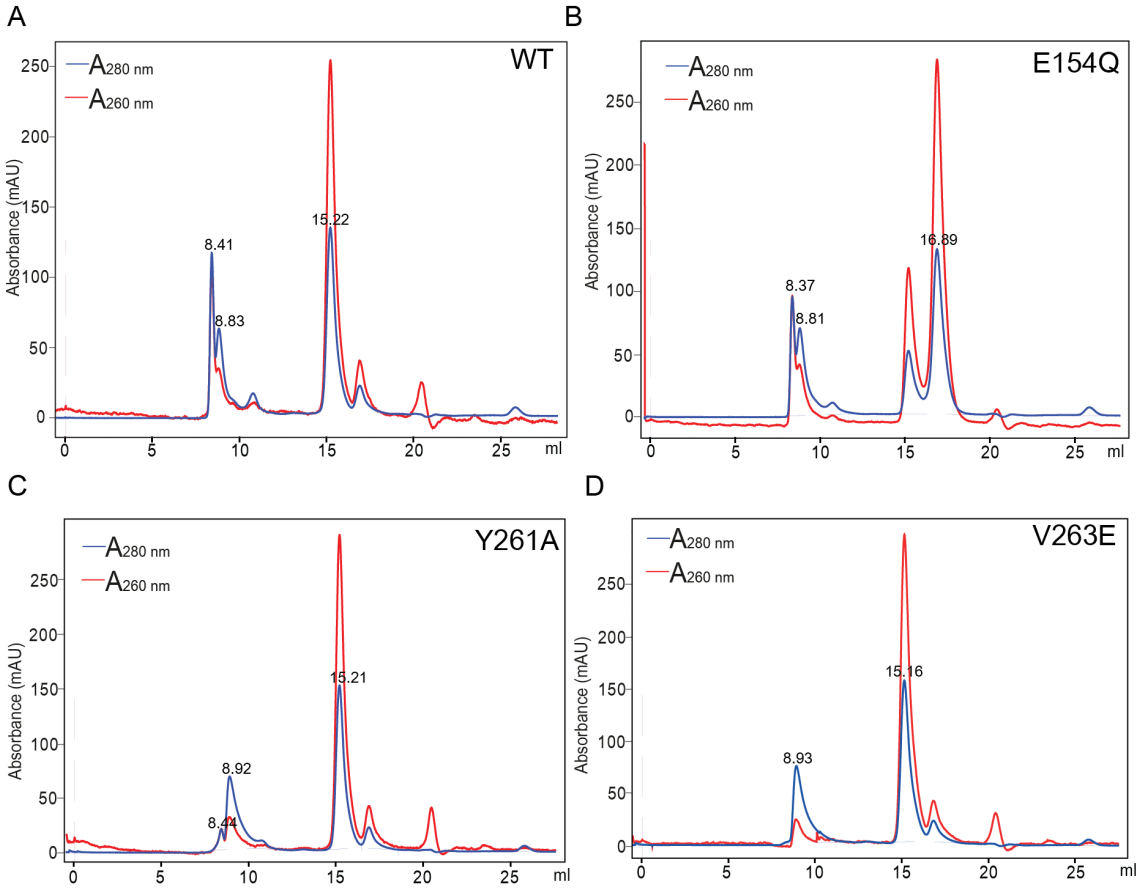

Figure 3-11 MVV intasome assembly assay.

Analysis of MVV intasome assembly reaction using (A) WT, (B) E154Q, (C) Y261A, or (D) V263E IN by size exclusion chromatography. The complex was assembled and purified in the absence of Mg^{2+}. Absorbance of proteins at 280 nm and DNA at 260 nm indicated in blue and red, respectively. The elution volumes of the intasome complex (~8.4 ml), proteins (~8.85 ml) and vDNA are indicated on the peaks.

To detect the intasome in the presence of co-eluting excess protein, vDNA was labelled with Cy3 fluorescent dye. Chromatography was performed while recording absorbance at 550 nm, corresponding to the peak absorbance of Cy3 (Figure 3-12A). The specific absorption of the Cy3 tagged DNA at 550 nm displays two peaks, the intasome peak between 8.2 ml to 8.5 ml, without the free protein peak, and the free vDNA peak around 15 ml.

WT (~38 mAu), E154Q (~25 mAu), R231E (~8 mAu), QQQ211PGG (~8 mAu) and Y261A (~5 mAu) readily displayed detectable intasome peaks, while Y261E displayed minor intasome peak of around 3 mAu (Figure 3-12B and C). No detectable nucleoprotein complex formation was observed with the other partially active mutants, F223A and V263E. The remaining mutants did not form intasomes.

Thus, compromised intasome assembly can explain the greatly reduced strand transfer activities of the MVV IN mutants studied in this work. R231E behaved similar to WT in multimerisation (Figure 3-6) however, in agreement with the involvement of Arg-231 in interactions with both vDNA and target DNA (Ballandras-Colas et al., 2017), the intasome assembly of R231E was greatly affected (Figure 3-12B). This led to its strongly reduced ability to perform strand transfer *in vitro* (Figure 3-9 and 3-10). The multimerisation of Y261E and QQQ211PGG was not detected as tetramers in solution, while Y261A possibly assembled into tetramers (at 8 mg/ml), but these mutants displayed strand transfer activities and assembled into intasomes *in vitro*. Although F223A and V263E displayed weakly detectable levels of strand transfer activities, the intasome assembly was not detected *in vitro*.

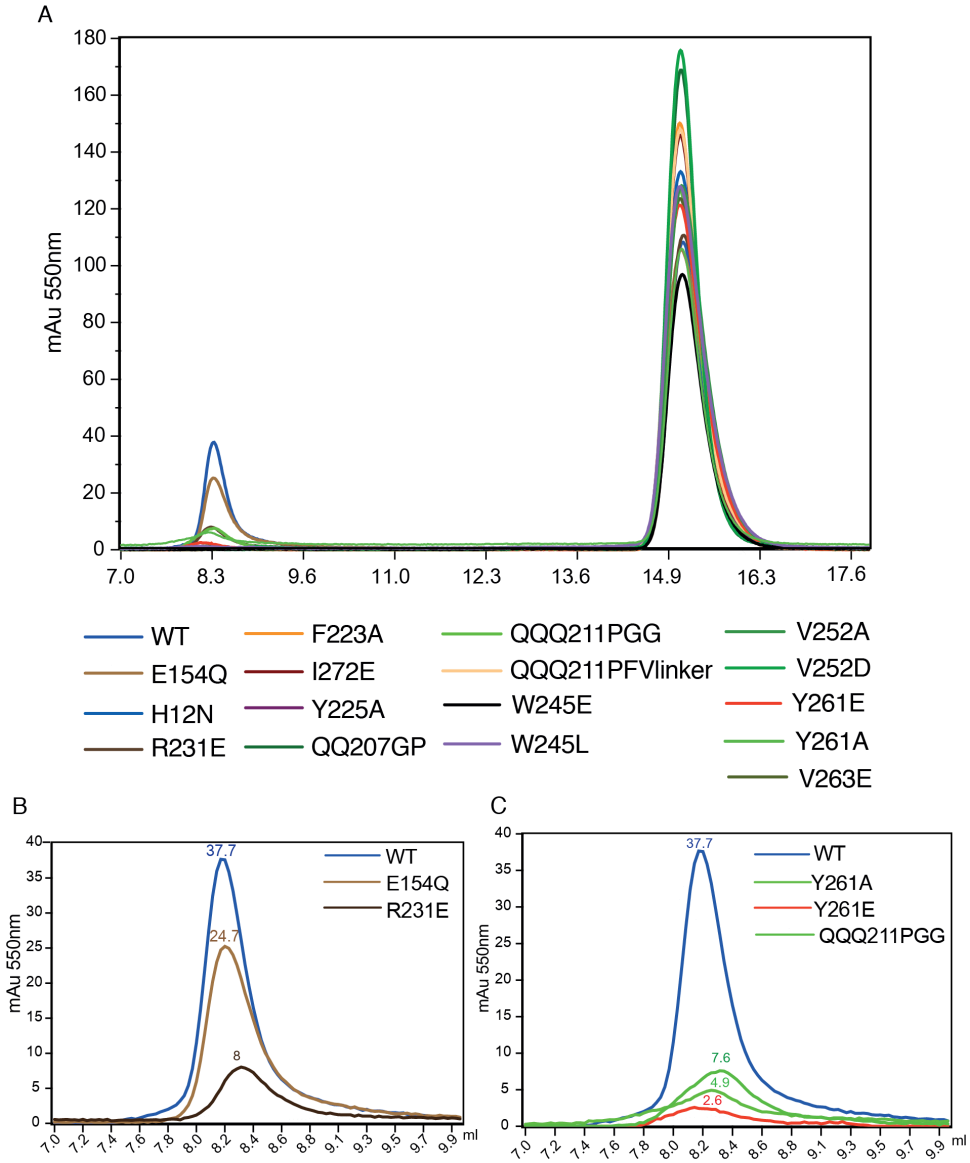

Figure 3-12 MVV intasome assembly with Cy-labelled vDNA

(A) Chromatograms (elution volumes 7 to 17.6 ml) of the MVV intasomes assembled with Cy3-tagged vDNA. Absorbance was measured at 550 nm, with only peaks specific for vDNA containing complex (intasome), eluting ~8.4 ml and excess vDNA, eluting at ~15 ml observed. Chromatograms (elution volumes 7 to 9 ml) of the MVV WT intasomes with (A) E154Q and R231E, and (B) Y261A, Y261E and QQQ211PGG. The absorption units at 550 nm (mAu 550 nm) for intasome peaks detected, WT (37.7 mAu), E154Q (24.7 mAu), R231E (8 mAu), QQQ211PGG (7.6 mAu), Y261A (4.9 mAu) and Y261E (2.6 mAu) are indicated on the graphs.

3.7 Conclusions

The importance of the hexadecameric assembly of the MVV intasome was tested by a radical amino acid substitution strategy targeting its unique features. Ten single amino acid substitutions were introduced to interfere with the inter- and intra-tetramer interactions in the CTD bridge. Mutations destabilising the α-helical CCD-CTD linker were also designed. These targeted two separate Gln repeats at IN positions 207, 208 and 211 to 213, and inserting 11 amino acid residues from PFV CCD-CTD linker.

All the amino acid substituted proteins expressed in bacterial cells, purified with similar chromatogram profiles to the WT recombinant protein (Figure 3-2 and 3-3), suggesting no significant effects on the protein stability or folding. Relatively high quantities of all the mutant forms of the protein were purified to a desirable protein concentration, ~30 mg/ml, with the exception of the tetramerisation defective control mutant, H12N and the intra-tetramer interaction targeting V252A mutant, which could only be purified and concentrated in smaller quantities.

In solution, WT, active site mutant E154Q, and R231E, displayed similar multimerization properties, existing in a tetramer-octamer equilibrium. However, most of the other amino acid substituted proteins tested, including H12N, were generally defective for tetramerisation and failed to assemble into higher order multimers (Figure 3-6). V252A assembled into tetramers in solution. At the highest concentration of injection, 8 mg/ml, W245E and W245A behaved similarly by assembling into higher order multimers and Y261A and W245L formed tetramers.

The majority of IN proteins were defective for strand transfer activity *in vitro* (Figure 3-9 and 3-10). Five of the mutant INs displayed detectable levels of strand transfer activities. F223A and QQQ211PGG only showed FS integration products at the higher IN concentration tested of 1.1 μM. The mutants that were defective for strand transfer activity failed to assemble into intasomes, while some of the partially active mutants assembled into intasomes with reduced capabilities (Figure 3-9, 3-12).

Although the lentiviral CCD-CTD linker containing 20 amino acid residues is not well conserved, its α-helical configuration appears intact within the genus (Ballandras-Colas et al., 2017). Destabilising this α-helix by substituting the Gln stretches led to partial or complete loss of strand transfer activity *in vitro*, due to these mutants'

inability to multimerise and assemble into stable intasomes. The attempted increase in the length of the CCD-CTD linker, by replacing residues 211 to 213 with eleven amino acid residues from the PFV IN CCD-CTD linker, not only led to the formation of dimers in solution, and consistently, this protein also lost its ability to assemble into a functionally active intasome. Overall, there is excellent correspondence between the ability of the MVV IN to form multimers, to assemble into intasomes and carry out strand transfer *in vitro*.

The data presented above highlights the importance of hexadecameric assembly of the MVV intasome for strand transfer activities. The following section (Chapter 4) further analyses the importance of this extended intasome architecture in MVV viral infectivity, by testing the mutants which displayed detectable strand transfer activity *in vitro* for their abilities to support MVV infectivity. In Chapters 5 and 6, functional consequences of the lentiviral intasome assembly are tested by exploring the necessity of LEDGF/p75 for the intasome assembly and its stability.

Chapter 4. Testing the importance of the hexadecameric intasome assembly in the virus

4.1 Aims

The amino acid substitutions introduced into MVV IN to disrupt the unique features of the lentiviral intasome structure led to a number of important observations described in Chapter 3. In particular, some of the IN variants that were generated had reduced abilities to multimerise into tetramers, and strand transfer activity was affected, presumably due to defects in assembling into intasomes *in vitro*. To validate these findings, the importance of the CTD-CTD interface interactions as well as the α-helical nature of the CCD-CTD linker were studied in the context of the viral infection. To this end, selected mutations were introduced into an MVV vector packaging construct.

4.2 MVV-based single cycle lentiviral vector system

In order to investigate the effects of the amino acid substitutions in MVV IN, in the context of viral infection, a single-cycle MVV viral system developed in the laboratory of Dr. David Griffiths (Moredun Research Institute, UK) (McLean and Griffiths, Manuscript in preparation) was used. The system comprises four plasmids, pCAG-MV-GagPol-CTEx2 (the packaging construct encoding MVV Gag-Pol), pCVW-CG-GFP (reporter plasmid, expressing GFP), pCV-VMV-Rev (encoding MVV Rev), and pMD2.G (a vector expressing VSV-G envelope glycoprotein). Details of the plasmids are listed in Table 4-1. Dr. Griffths additionally provided pCAG-MV-GagPol-CTEx2-IN$^{D66A/E154A}$, which is an IN active site mutant form of the packaging plasmid, carrying point mutations inactivating the IN active sites.

In order to facilitate mutagenesis, pCAG-MV-GagPol-CTEx2 was modified by introducing restriction sites flanking the IN-coding region. The original construct harbours a pair of unique restriction sites: *Pas*I (5′-CCCWGGG-3′, 115 bp upstream of IN region, within the RT coding region) and *Dra*III (5′-CACNNNGTG-3′, 587 bp downstream of the second CTE element) (Figure 4-1A). First, the 2.5-kb *Pas*I/*Dra*III fragment was subcloned into pBluescript SK(-) bacterial vector (obtained from Dr.

Rebecca Newman, Pavel Tolar Lab, The Francis Crick Institute). Two silent point mutations (changing the nucleotide sequence ACAGGA at positions 5569-5574 to ACCGGT) were introduced to create an *Age*I restriction site, immediately upstream of the IN coding region. Next, an *Xho*I restriction site (CTCGAG, positions 6427-6432) was introduced between the Gag-Pol stop codon and the first CTE element (Figure 4-1B). The original *Pas*I/*Dra*III fragment in pCAG-MV-GagPol-CTEx2 was then replaced with the mutant version to obtain pCAG-MV-GagPol-CTEx2-IN-Age/XhoI. The modification allowed seamless replacement of the IN coding region with the variants used in this work.

Table 4-1 MVV single cycle vector system

Plasmid	Function
pCAG-MV-Gagpol-CTEx2	Visna Gag-pol expression plasmid
pCAG-MV-GagPol-CTEx2-IN$^{D66A/E154A}$	Visna Gag-pol expression plasmid, with IN mutations D66A and E154A
pCAG-MV-GagPol-CTEx2-IN.KV1772-Age/XhoI	Visna Gag-pol expression plasmid, modified with insertions of AgeI and XhoI restriction sites flanking IN
pCVW-CG-GFP	Visna vector plasmid with CMV-GFP-WPRE expression cassette
pCVW-CG-Luc	Visna vector plasmid with CMV-Luc-WPRE expression cassette
pMD2.G	VSV-G expression (from Addgene)
pCMV-VMV-Rev	Codon-optimised Visna Rev expression plasmid

The recombinant MVV IN constructs used in this study were based on MVV isolate KV1772. However, the original MVV packaging construct, pCAG-MV-GagPol-CTEx2, was built based on the isolate MV1 that shares 86% of IN amino acid sequence identity with KV1772. Therefore, the IN-coding region of the packaging construct was replaced with that of KV1772 IN, and the modified plasmid (pCAG-MV-GagPol-CTEx2-IN.KV1772-Age/XhoI). For clarity, the viral particles produced using this modified construct are referred to as WT-KV1772, while those produced using the original pCAG-MV-GagPol-CTEx2 packaging construct are referred to as WT-MV1. Next, the original IN coding region was replaced with the mutant forms. The mutations targeting the CTD-CTD interface that displayed *in vitro* strand transfer activity (Figure 3-9), Y261A, V263E, and F223A, as well as the CCD-CTD linker mutation QQQ211PGG, were introduced into pCAG-MV-GagPol-CTEx2-

IN.KV1772-Age/XhoI. In addition, two control mutants H12N and E154Q were constructed.

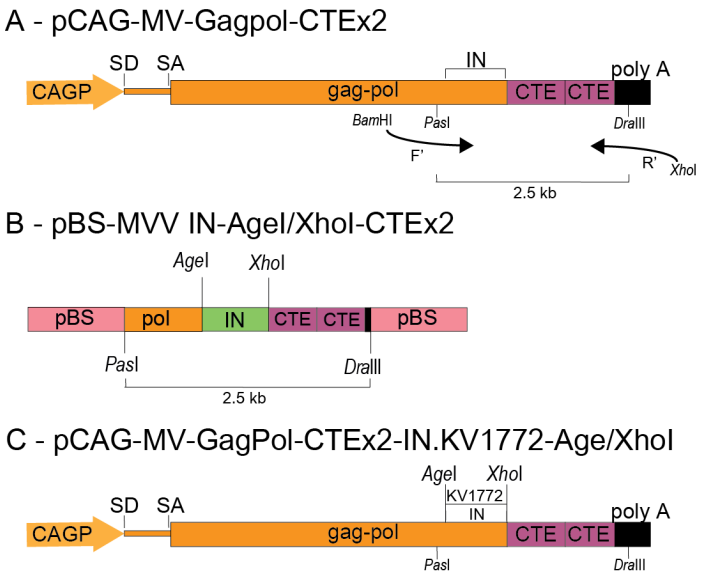

Figure 4-1 Modifications of the MVV packaging construct.

(A) The 2.5-kb region of pCAG-MV-Gagpol-CTEx2 between *Pas*I and *Dra*III sites was amplified by PCR. (B) The PCR fragment was ligated into restriction-digested pBluescript SK(-) (pBS). PCR-based modifications were made to flank the IN-coding region with *Age*I and *Xho*I and replace the coding region of WT-MV1 IN to that of WT-KV1772. (C) The modified 2.5-kb fragment was then inserted into the original MVV packaging construct, to generate pCAG-MV-GagPol-CTEx2-IN.KV1772-Age/XhoI.

4.2.1 Production of MVV single cycle viruses

MVV WT-KV1772 and mutant viruses were produced by transfection of plasmid DNA into HEK-293T cells using PEI. To obtain sufficient virus titres, large-scale transfections were performed using at least six 15-cm dishes per virus variant, detailed in section 2.4.1. Viruses were harvested 48 h post-transfection, concentrated by ultracentrifugation and stored -80ºC.

4.2.2 Quantification of viruses using Taq-PERT

All the viruses were produced under similar conditions, however as the mutations created in the *gag-pol* gene may impact the total viral titres (Lu et al., 2005a), they were quantified by Taq-PERT (Vermeire et al., 2012, Li et al., 2015). Briefly, 5 µl viral stock was lysed to release the RT enzyme. MS2 phage RNA was provided as a template, along with appropriate primers and dNTPs to allow reverse transcription. MS2 cDNA synthesised was amplified in a PCR reaction with 40 cycles using TaqMan real-time qPCR. A standard curve was generated with known quantities of recombinant HIV-1 RT, which allowed for the relative quantification of the viral titres. Further details of the viral titre quantification are described in section 2.6.1.

4.2.3 Western blotting

Introducing mutations in the IN-coding region of *gag-pol* may lead to aberrant viral morphogenesis, either due to defects in incorporation or proteolytic processing of Gag and Gag-Pol polyproteins (Shin et al., 1994, Engelman et al., 1995, Bukovsky and Göttlinger, 1996, Lu et al., 2005a). Such mutations may render the virus non-infectious by terminating the infection process prior to reverse transcription and integration. The IN-coding region in the MVV Gag-Pol plasmid was also modified by introducing restriction sites, *Age*I and *Xho*I, as well as mutations designed to create amino acid substitutions in IN. Gag processing in the viral particles was tested by Western blotting of viral lysates using polyclonal antibody to MVV capsid/p24 protein. Viral lysates normalised for RT activity were analysed. To enable quantification, dilutions of WT-KV1772 vector lysate were tested in the same experiment.

The dilutions of WT-KV1772 virus displayed a linear decrease in the intensity of the p24 band. The intensity of the p24 band in samples of E154Q, V263E, H12N and Y261A viruses was similar to that of WT-KV1772 vector (Figure 4-2). Although equal viral quantities were loaded, QQQ211PGG and F223A had noticeably reduced p24 band intensities than 100% WT-KV1772, with QQQ211PGG being 30% of WT-KV1772 and F223A being 67% (Figure 4-2).

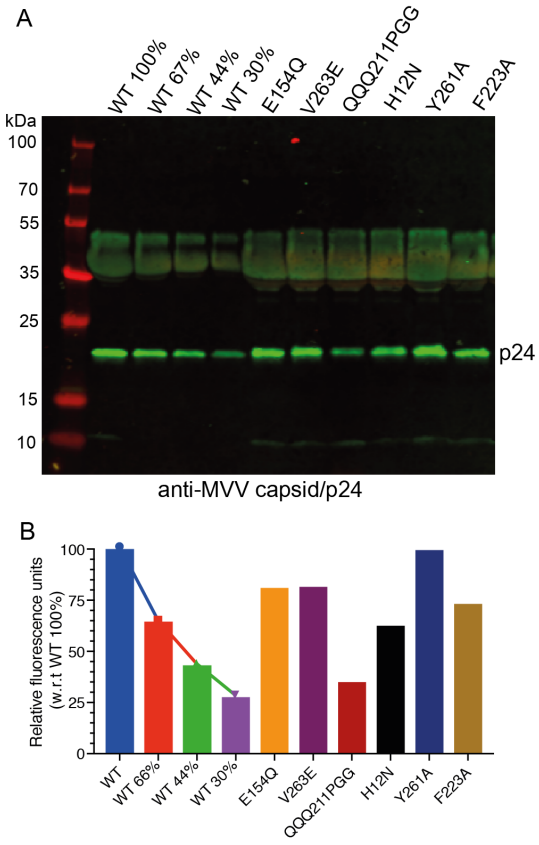

Figure 4-2 Capsid/p24 Western blotting of MVV virus.

(A) 0.05 mU RT-normalised mutant viruses, along with indicated dilutions of WT - 0.05 mU (100%), 0.0335 mU (67%), 0.022 mU (44%) and 0.015 mU (30%) were tested for CA/p24 content, by Western blotting using anti-CA/p24 polyclonal antibodies. Detection was carried out using near-infrared fluorescence of appropriately conjugated secondary antibodies. (B) The fluorescence units measured were normalised to the values of WT at 0.05 mU RT (set to 100%) and plotted as a bar graph. The reduction in fluorescence intensity in WT dilutions is indicated by a line.

4.3 Initial infectivity analysis

The MVV packaging construct pCAG-MV-GagPol-CTEx2 from the single cycle system was modified by replacing the IN-coding region with that of KV1772 isolate. It was important to verify that this did not compromise infectivity of the lentiviral vector. To this end, infection experiments were carried out in HEK-293T cells with (i) WT-KV1772, (ii) WT-MV1, and (iii) $IN^{D66A/E154A}$ viruses using 350 µl and 150 µl of each viral stock. Infected cells were harvested 4 d and 6 d post-infection, fixed with paraformaldehyde and analysed by flow cytometry to quantify GFP-positive cells. Cell debris were gated out of the live cell population, and single cells were separated from doublets. Further on, 10,000 single cells were analysed for GFP expression. The gating strategy for GFP-positive cells was based on non-infected HEK-293T cells, used as a negative control in these experiments (Figure 4-3).

Fraction of GFP-positive cells was measured 4 d and 6 d post-infection and normalised to WT-KV1772 at 4 d post-infection with 350 µl inoculation. At the higher viral volume inoculation (350 µl), infectivity of WT-KV1772 and WT-MV1 viruses was around 100%, while infectivity with the IN active site double mutant was ~8% (Figure 4-4). A similar pattern was observed at the lower viral volume inoculation (150 µl), with WT-KV1772 infectivity measured around 80% and $IN^{D66A/E154A}$ being 5% (Figure 4-4). At the later time point, 6 d post-infection, the WT-KV1772 infectivity was reduced by around 5 to 10%, while the infectivity of the IN active site double mutant remained constant (Figure 4-4). The preliminary infection data indicated that the manipulations made to generate the WT-KV1772, including the introduction of the silent mutations and the replacement of the IN-coding region in the packaging construct did not cause dramatic infectivity defects. Subsequently, 350 µl volume of the virus used was approximated as 16.8 mU of RT by Taq-PERT.

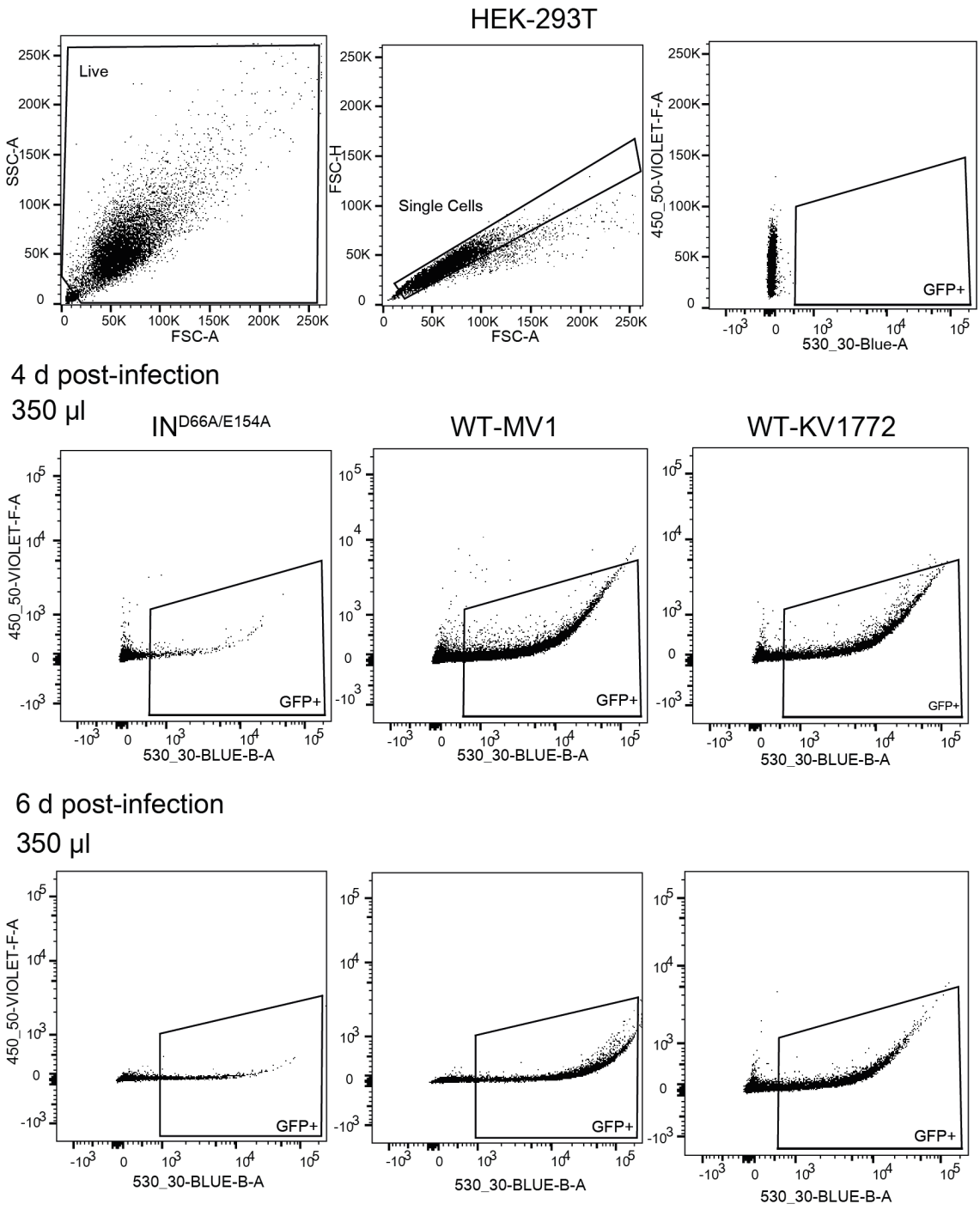

Figure 4-3 Analysis of GFP expression in infected cells by flow cytometry.

The gating strategy for infected cells, with non-infected HEK-293T cells used as GFP-negative control. GFP-positive cells analysed from single cell populations of infected cells at 4 d and 6 d post-infection time points, data shown for cells inoculated with 350 µl virus.

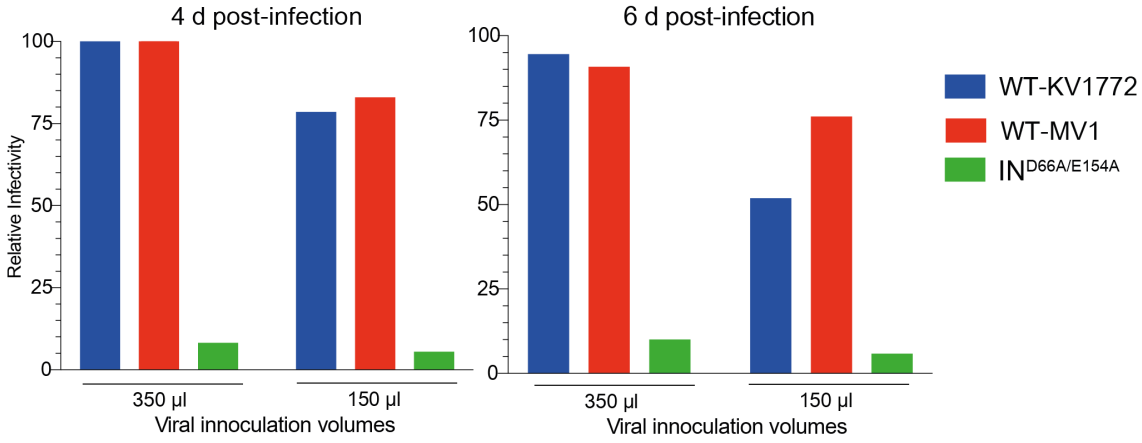

Figure 4-4 Infectivity of WT-MV1 and WT-KV1772 vectors.

Infection of HEK-293T cells with MVV vectors, WT-KV1772, WT-MV1 and IN active site mutant, $IN^{D66A/E154A}$ was performed once, and numbers of GFP-positive cells were measured by flow cytometry. The fractions of GFP-positive cells were normalised to that of WT-KV1772 with 350 µl inoculation at 4 d post-infection (64.5% cells scored GFP-positive under this condition).

In order to optimise the infectivity titres and timelines for infections, HEK-293T cells were infected with 16.8, 8.4 and 4.2 mU RT-normalised WT-KV1772, E154Q and H12N viruses (Figures 4-5 and 4-6). Infected cells were harvested 4 d and 7 d post-infection and fixed in paraformaldehyde, before being analysed by flow cytometry for GFP expression. Flow cytometry analysis was carried out by gating cell populations for live cells, single cells and GFP positive cells. Non-infected HEK-293T cells were used as GFP-negative controls.

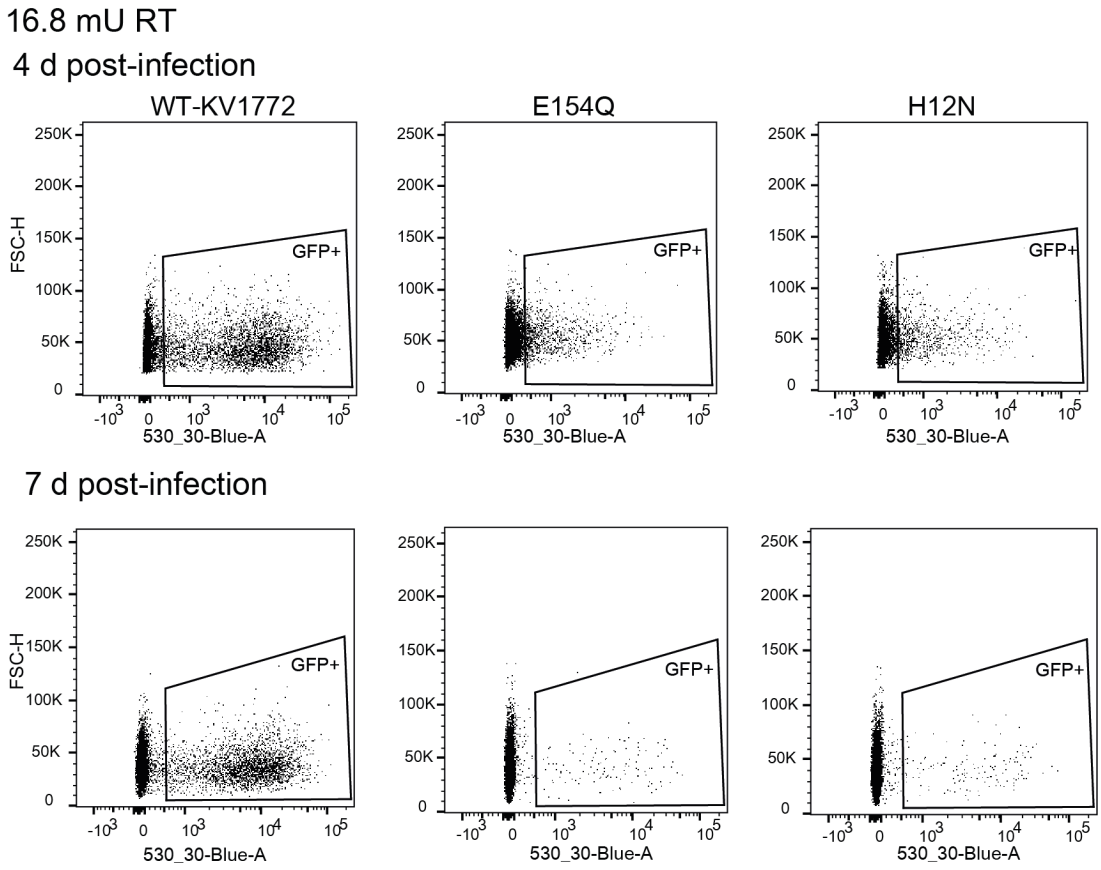

Figure 4-5 GFP analysis of infected cells by flow cytometry.

The gating strategy for cells infected with MVV WT-KV1772, E154Q and H12N viral vectors and harvested 4 d and 6 d post-infection. GFP-positive cells analysed from single cell populations of infected cells; data shown for 16.8-mU RT viral inoculation.

At the earlier time point of 4 d post-infection, the control mutants were approximately half as infectious as WT-KV1772. At the later time point of 7 d post-infection, WT infectivity reduced by ~25% and the control mutants infectivity was under 5%. The infectivity of the MVV controls measured 7 d post-infection, was similar to that of $IN^{D66A/E154A}$ measured 6 d post-infection (Figures 4-4 and 4-6). The steep reduction in GFP positive cell counts was expected for integration-defective mutants, as non-integrated vDNA forms, competent for GFP expression, are diluted and/or degraded with cell divisions.

Generating large quantities of mutant forms to analyse their infectivity by GFP measurement in multiple experiments proved to be very challenging. Higher mU of RT-normalised virus was necessary to reliably detect and gate GFP-positive cells. In order to overcome this limitation, the *gfp* reporter gene was replaced with *luc* gene encoding firefly luciferase, which allowed for very sensitive quantification. The modified transfer vector pCVW-CG-Luc was constructed and kindly provided by Dr. Wen Li (Alan Engelman's laboratory, Dana-Farber Cancer Institute, Boston, MA, USA).

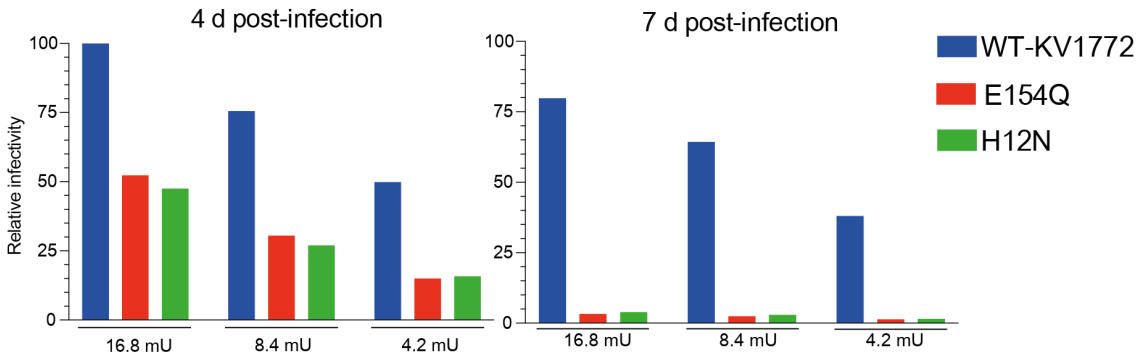

Figure 4-6 Initial infectivity analyses.

HEK-293T cells were infected with 16.8 mU, 8.4 mU and 4.2 mU RT-normalised MVV viral vectors, WT-KV1772, E154Q and H12N, and the fractions of GFP-positive cells were measured by flow cytometry. Infected cells were harvested 4 and 7 d post-infection and percentage of GFP-positive cells was normalised to that of WT-KV1772, measured 4 d post-infection with 16.8 mU RT (40% cells scored GFP-positive under this condition).

4.4 Infectivity analyses of MVV mutants

4.4.1 Optimisation of HEK-293T infection with the MVV vectors

HEK-293T cells were seeded into 48-well plates and infected with equal virus amounts corresponding to 1, 0.5 and 0.25 mU RT activity. 2 d post-infection, cells were expanded into 6-well dishes. After retroviral infection, unintegrated vDNA present in the cells is only cleared after multiple cell divisions (Butler et al., 2001). In order to specifically obtain signals from the protein synthesised from integrated proviral DNA, infectivity was measured 7 d post-infection. To this end, infected cells were gently lysed and their protein contents were collected. D-luciferin was provided as a substrate for luciferase. The enzyme converts this substrate to oxyluciferin, emitting a photon as by-product (Figure 4-7). Luciferase levels were measured in relative light units (RLUs) using a luminometer and normalised to the total protein content of cell extracts, as detailed in section 2.6.3.

Luciferase levels in cells infected at 1 and 0.5 mU RT with WT was 10 million, which reduced to 8.5 million at 0.25 mU RT. Both the control mutants were ~100-fold defective for integration 7 d post-infection (Figure 4-8). Based on these observations, subsequent infections were carried out with 0.5 mU RT-normalised virus.

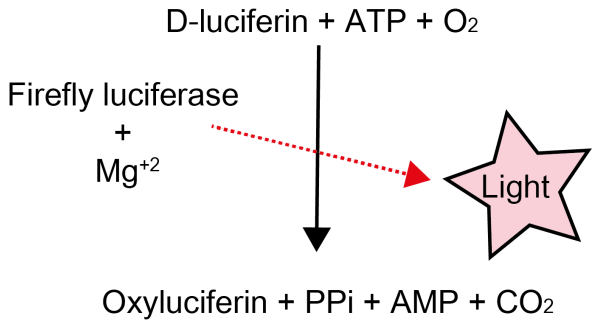

Figure 4-7 Luciferase reaction.

Firefly luciferase, synthesised in infected cells, converts D-luciferin to oxyluciferin, generating luminescence. The amount of light produced is a measure of the infectivity of the virus.

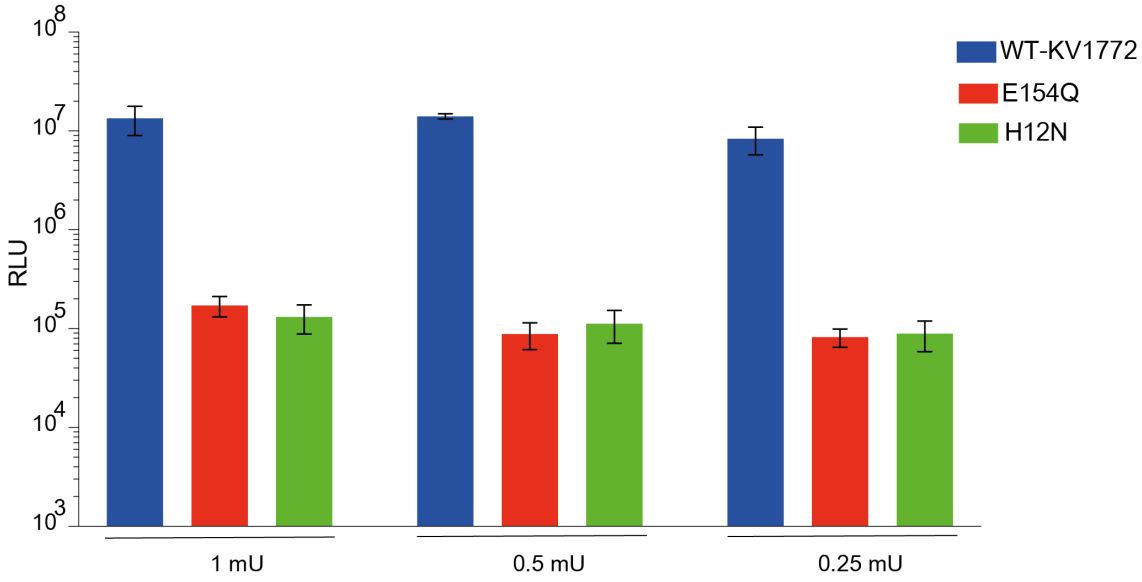

Figure 4-8 Infectivity titre optimisation.

Luciferase activity was measured 7 d post-infection from cells infected with 1 mU, 0.5 mU and 0.25 mU RT-normalised WT-KV1772, E154Q and H12N viral vectors. Luciferase units were normalised to the protein concentration in the cell lysates to obtain relative luciferase units (RLU). Means plotted of one infection with two technical repeats, with SD.

HEK-293T cells were seeded into 48-well plates and infected with equal amounts, 0.5 mU RT-normalised virus. 2 d post-infection, cells were expanded into 6-well dishes and analysed 7 d post-infection. Infectivity of the viruses was normalised with respect to that of WT-KV1772, which was considered 100%. As expected, E154Q and H12N were non-infectious, with luciferase readings <1%. All mutants tested, in spite of being partially active for strand transfer activity *in vitro* (Figures 3-9 and 3-10), were significantly less infectious than the WT-KV1772, behaving in a similar fashion to the control mutants, consistent with severe integration defects (Figure 4-9).

The IN mutations introduced to disrupt the MVV IN CTD-CTD interface interactions and to destabilise the α-helical CCD-CTD linker change residues located outside the CIC. All the observations reported here indicate the necessity and importance of these features not only for the multimerisation of INs (Figure 3-6) and their strand

transfer activities (Figures 3.9 and 3.10) *in vitro*, but for also for the viral infectivity. These results indicate that the mutations affect the formation and stability of the intasome both *in vitro* (Figures 3.12) and in the context of a virus, thus supporting the importance of the hexadecameric MVV intasome assembly observed in the cryo-EM structure (Ballandras-Colas et al., 2017).

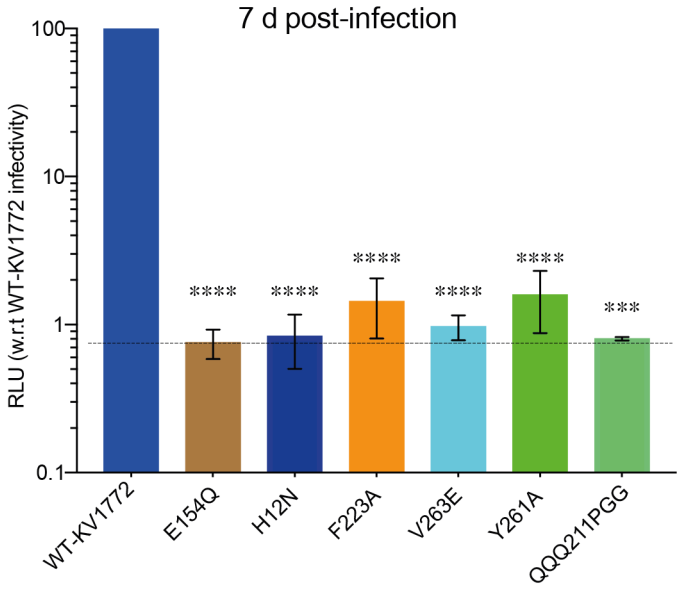

Figure 4-9 Infectivity of MVV-KV1772 vector mutants.

Luciferase activity was measured 7 d post-infection from cells infected with 0.5 mU RT-normalised MVV WT-KV1772, E154Q, H12N, F223A, V263E, Y261A and QQQ211PGG vectors. Infectivity of the mutants was normalised to the infectivity of WT-KV1772. Means with standard error of the mean plotted of three biological repeats, with two technical repeats each; Two-tailed Paired t tests (individual comparisons), $p<0.0001$ (****).

4.5 Dissection of the phenotypes of the mutant viruses

Alongside catalysing the integration reaction, IN has been implicated in playing a key role during other stages of the viral lifecycle. Deleting the IN coding region or introducing mutations to generate amino acid substitutions in this viral component were shown to cause pleiotropic defects in HIV-1 replication, especially during viral reverse transcription or/and virion morphogenesis (Shin et al., 1994, Taddeo et al., 1994, Engelman et al., 1995, Bukovsky and Göttlinger, 1996, Wu et al., 1999, Hehl et al., 2004, Fontana et al., 2015, Tekeste et al., 2015). Collectively such phenotypes are characterized by a reverse transcription defect, and the mutants sharing this property are termed as Class II IN mutants (Engelman, 1999), as elaborated in section 1.8.

Most amino acid substitutions introduced in this work were designed to target features outside of the CIC. Yet, the mutants that were partially active for IN strand transfer activity mutant *in vitro* (Figures 3-9 and 3-10), were non-infectious (Figure 4-9). In order to assess whether these mutants were defective specifically at the integration step, the viral ability to perform reverse transcription was first evaluated.

4.5.1 vDNA synthesis

Mutations introduced in IN can cause defects in the viral ability to perform reverse transcription (Shin et al., 1994, Taddeo et al., 1994). The majority of vDNA synthesis by reverse transcription is generally completed 6 h to 8 h post-infection. The vDNA synthesised is rapidly degraded by the cells within 24 h post-infection (Butler et al., 2002, Yoder et al., 2006). Under normal conditions, a fraction of the non-degraded vDNA is permanently inserted into the host genome at the step of integration.

4.5.1.1 Initial quantification of vDNA by intracellular reverse transcription

The vDNA synthesised as a consequence of reverse transcription was measured using quantitative PCR (qPCR). HEK-293T cells were seeded in 12-well plates and infected with DNase I-treated MVV virus particles at a titre of 0.3 mU RT. Infected cells were harvested 2 h, 8 h and 24 h post-infection, and DNA was extracted. A TaqMan probe based real-time qPCR was performed with primers targeting the *luc*

gene sequence (Section 2.6.4). The pCVW-CG-Luc plasmid was used to generate a standard curve, to determine the relative copy numbers of the vDNA present in the infected cells.

In a control condition, cells were infected with WT-KV1772 vector in the presence of 150 μM azidothymidine (AZT, 3′-azido-2′,3′-dideoxythymidine). AZT is a clinical HIV-1 nucleoside RT inhibitor (NRTI) (Rai et al., 2018). This small molecule acts as a chain terminator mimicking the native nucleotide (thymidine) and displays activity against a wide range of retroviruses, including MVV (Thormar et al., 1993). Importantly, to act as a chain terminator, AZT must to be phosphorylated into its triphosphate form by cellular nucleotide kinases (Furman et al., 1986), and thus needs to be present prior to infection.

WT-KV1772 vDNA copy number 8 h post-infection was normalised to 100% and the copy numbers of the control mutants plotted with respect to WT-KV1772 at 8 h post-infection. The vDNA quantities observed at the measured time points were in accordance to previously published data with HIV-1 (Butler et al., 2002, Yoder et al., 2006). Wherein 2 h post-infection, only a minor quantity of vDNA was synthesised, with a majority of the vDNA synthesised 8 h post-infection. As expected, at the later time point of 24 h post-infection, a dramatic reduction in vDNA levels was observed, presumably due to degradation of a large fraction of the reverse transcription products in infected cells (Figure 4-10).

Unexpectedly, E154Q, the IN active site mutant, was reproducibly 2.5 times more competent for vDNA synthesis than the WT-KV1772. Furthermore, the tetramerisation defective H12N mutant was also highly compensated, 3.5-fold more competent than WT-KV1772 for vDNA synthesis. Both of these observations are in stark contrast with the corresponding HIV-1 phenotypes. Indeed, HIV-1 IN E152Q (HIV-1 IN Glu-152 is equivalent to MVV IN Glu-154) is a canonical class I mutant with normal reverse transcription (Lu et al., 2005b), while HIV-1 H12N displayed a severe class II phenotype, with a strong defect in vDNA synthesis (Wu et al., 1999).

Importantly, a carryover of pCVW-CG-Luc plasmid, which was used to produce viral particles, could affect the vDNA measurements. However, vDNA levels measured at 2 h post-infection were considerably lower than those during the burst synthesis time point of 8 h post-infection, indicating that most of the observed signal was from vDNA

produced by reverse transcription. Moreover, as expected, AZT greatly suppressed vDNA synthesis (Figure 4-10). However, given the unanticipated findings, including the increased vDNA synthesis by the H12N and E154Q controls, an alternative method to quantify proviral DNA was developed, which would not be affected by a potential carryover of plasmid DNA.

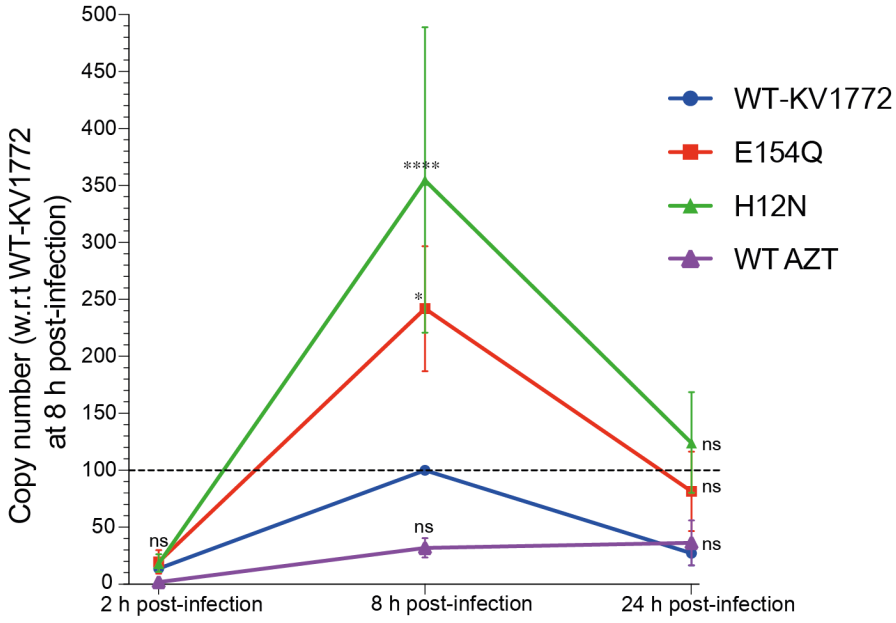

Figure 4-10 Quantification of viral vDNA synthesised in infected cells.

HEK-293T cells were infected with 0.3 mU RT-normalised WT-KV1772, E154Q and H12N viral vectors and harvested 2, 8 and 24 h post-infection. Viral cDNA was quantified by real-time qPCR, with primers annealing within the *luc* gene. As a negative control, the cells were infected with WT-KV1772 in the presence of 100 µM AZT (a nucleoside RT inhibitor). Quantities of vDNA were determined from a standard curve generated using pCVW-CG-Luc. The values were normalised to that for WT-KV1772 vDNA synthesis 8 h post-infection, which was set to 100%. Means values plotted with standard error of the mean; Two-tailed paired t-tests (individual comparisons) between WT-KV1772 and mutants at each time point. 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****).

4.5.1.2 Strategy to measure late MVV reverse transcription products

The final (late) product of reverse transcription is the full-length vDNA, flanked by LTRs at both 5′ and 3′ ends. By design, pCVW-CG-GFP and pCVW-CG-Luc vectors lack full-length 5′ LTR, relying on CMV promoter for expression of viral RNA during vector production stage (Figure 2-2). Thus, PCR amplification of the late reverse transcription (LRT) product may allow for specific quantification of vDNA even in the presence of contaminating plasmid DNA. The U3 region at 5′ end is added during reverse transcription. PCR primers to specifically amplify a DNA fragment spanning U3 and the 5′ region of MVV *gag* gene were designed to develop an assay to specifically measure vDNA (Figure 4-11A).

One of the challenges posed with adopting the late reverse transcription (LRT) qPCR assay was the lack of template DNA to generate standard curves. The vDNA products to be quantified are only synthesised in cells after infection and were not readily available as plasmids. Therefore, a plasmid was constructed to generate the standard curve for the qPCR by two-step splicing PCR method (Figure 4-11B). A similar strategy was used to introduce mutations into pCPH6P-MVV-IN plasmid, as described in sections 2.2.2, 2.3.3.

In the first step, two PCR reactions were carried out, using an outer F′ primer (PC1803, binds pCVW-CG-Luc at position 4803 to 4823) with an inner R′ primer (PC1804, binds pCVW-CG-Luc at position 5206 to 5186), and an inner F′ primer (PC1805, binds pCVW-CG-Luc at position 711 to 731) with an outer R′ primer (PC1806, binds pCVW-CG-Luc at position 1138 to 1118), as illustrated (Figure 4-11B). The PCR products were then amplified together in a reaction in the presence of the outer primers and ligated into pCR4-TOPO (Invitrogen). The resultant construct, pCR4-TOPO-MVV-LRT, was confirmed by sequencing analysis and used to generate standard curves to quantify LRT products in infected cells.

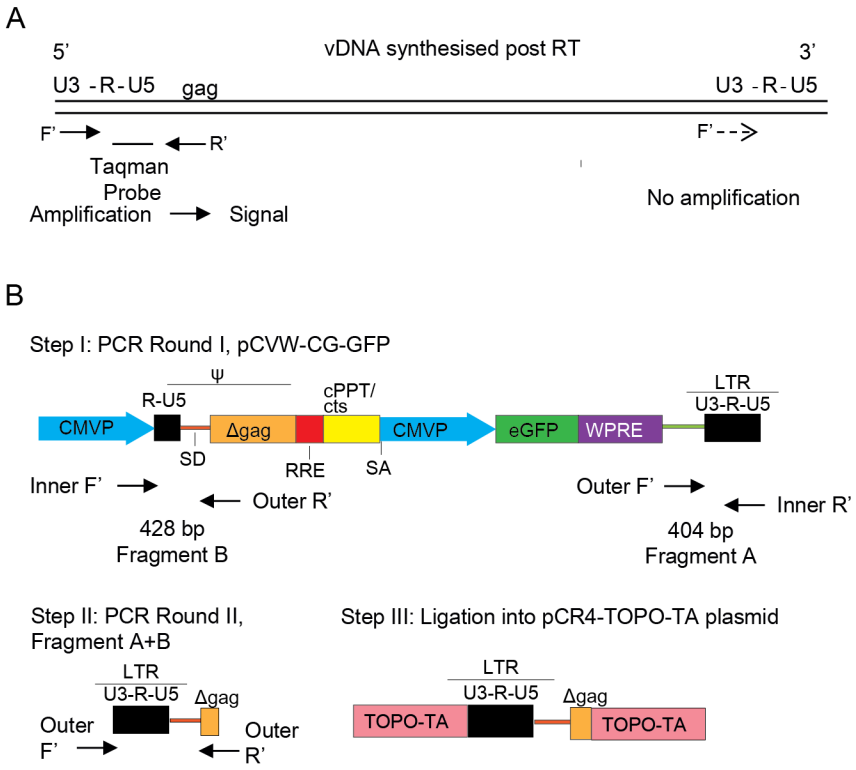

Figure 4-11 Strategy for the LRT product quantification.

LRT vDNA products contain U3 at the 5' *gag* region, which are not present in the transfected viral plasmids. (A) Primers designed to amplify the 5' U3-*gag* region of the vDNA. (B) Steps involved in the generation of the standard plasmid (pCR4-TOPO-MVV-LRT) for LRT qPCR. Step I: Two PCR reactions (PCR A and PCR B) were carried out using appropriate primers, amplifying the indicated regions of the pCVW-CG-GFP plasmid. Step II: The two PCR reaction products are mixed in equimolar ratios and amplified with only the outer primers, to generate the complete fragment of interest (U3 end of the LTR at the 5' end into the gag region). Step III: The generated fragment was cloned into the pCR4-TOPO-TA vector, to generate pCR4-TOPO-MVV-LRT plasmid.

Another technical issue with the experiments described in the previous section was incomplete inhibition of MVV vDNA synthesis by AZT (Figure 4.10). An alternative to AZT is 2′,3′-dideoxycytidine (ddC), which is also an NRTI, acting as a chain terminator once incorporated during DNA synthesis (Broder and Yarchoan, 1990).

Crucially, ddC was shown to be more effective at inhibiting MVV RT activity, compared to AZT (Thormar et al., 1993).

HEK-293T cells were seeded in 12-well plates and infected with DNase I-treated viruses at a titre of 0.3 mU RT per well. Infected cells were harvested 2 h, 8 h and 24 h post-infection and DNA was extracted. In a negative control condition, cells were infected with WT-KV177 vector in the presence of 100 μM ddC. A TaqMan probe based real-time qPCR was performed with primers targeting the U3 region of the LTR and *gag* (Section 2.6.5).

All measurements were normalised to WT-KV1772 LRT signal measured at 8 h post-infection. Predictably, 2 h post-infection (Figure 4-12), no synthesis of vDNA was detected. A burst of vDNA synthesis was observed 8 h post-infection, as in the previous set of experiments. Cells pre-treated with ddC and infected with WT-KV1772 were quantified to be around 22% and this has been set as the baseline for reverse transcription deficiency. 24 h post-infection, a drastic reduction in the total vDNA quantity in WT-KV1772 was detected, which is in agreement with the observations made with HIV-1 (Butler et al., 2002, Yoder et al., 2006). A similar pattern was observed with all the mutants tested.

At 8 h post-infection, E154Q, the enzyme active site mutant, behaved similar to WT-KV1772 (Figure 4-12). Surprisingly, even in this method of vDNA quantification, the tetramerisation defective H12N mutant was highly competent, five-fold higher than WT-KV1772 for reverse transcription, also observed in Figure 4-10. As stated in section 4.5.1.1, these observations are in disagreement with the previously reported HIV-1 reverse transcription data (Wu et al., 1999).

At 8 h post-infection, F223A behaved similar to WT-KV1772, while V263E and QQQ211PGG outperformed WT-KV1772 by ~1.5-fold, with their vDNA quantities being around 150% and 170%, respectively (Figure 4-12). Y261A was not as competent for reverse transcription, with its copy number being 65%. However, because Y261A outperformed WT-KV1772 supplemented with ddC, we conclude that even this mutant is at least partly competent for vDNA synthesis (Figure 4-12).

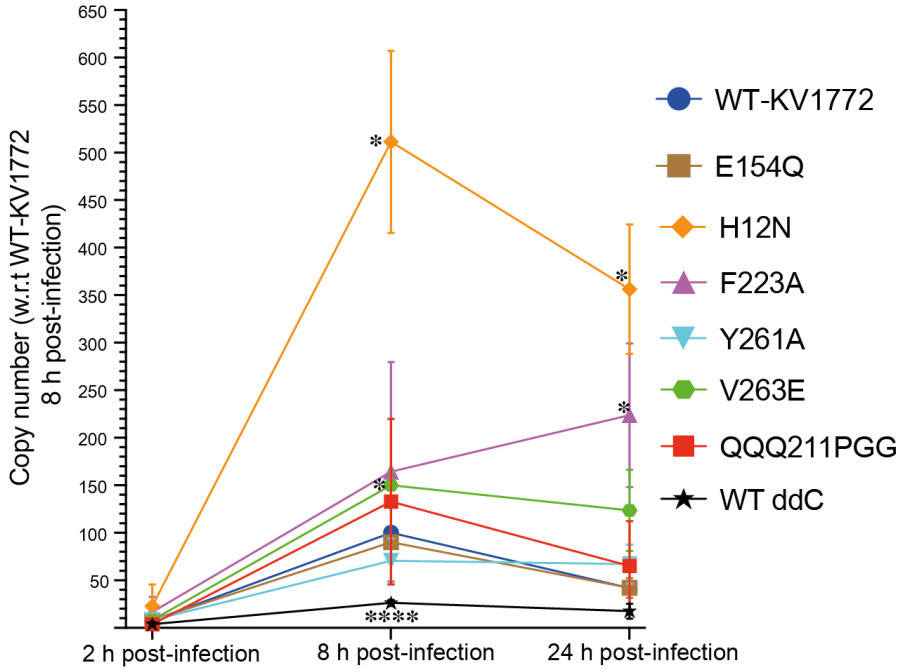

Figure 4-12 Quantification of LRT products in infected cells.

HEK-293T cells were infected with 0.3 mU RT-normalised MVV IN mutant viral vectors and harvested 2, 8 and 24 h post-infection. LRT vDNA was quantified by real-time qPCR, with primers annealing to the U3 and gag regions of the vector. Infection of WT-KV1772 in the presence of 100 μM ddC (a nucleoside RT inhibitor) served as a negative control condition. Quantities of the viral cDNA were determined from a standard curve generated using pCR4-TOPO-MVV-LRT plasmid. WT-KV1772 vDNA synthesis 8 h post-infection was normalised to 100%. Mean values from three biological repeats, plotted with SD; Two-tailed Paired t tests (individual comparisons) between WT-KV1772 and mutants at each time point. 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****). Only significantly different values were indicated, the remainder are ns.

4.5.2 Thin-section transmission electron microscopy of MVV virions

Amino acid substitutions in IN have been shown to cause defects in HIV-1 virion morphogenesis (Shin et al., 1994, Engelman et al., 1995, Fontana et al., 2015). A frequent HIV-1 morphogenesis defect associated with class II IN mutations manifests in failure to form mature capsid cores around the viral RNA genome, in some cases generating 'eccentric' forms of the virus (Fontana et al., 2015). As a consequence of this defect, viruses with the mutant IN forms were unable to undergo efficient reverse transcription. The amino acid substitutions within MVV IN CTD and CCD-CTD linker did not cause dramatic reverse transcription defects, and it can therefore be argued that they matured normally. However, to enable studies of potential class II IN mutants, a protocol to morphologically characterise MVV vector particles was optimised.

MVV virions isolated from infected sheep have been visualised by negative staining (Thormar and Cruickshank, 1965), as well as thin-section EM (TEM) (Weiland and Bruns, 1980, Gendelman et al., 1986), described in section 1.8. Although, no previous attempts were made to test the morphology of the MVV virions synthesised by transfecting plasmids in cells, MLV virions were synthesised by transient transfection of plasmids and visualised by TEM (Wight et al., 2014).

WT-KV1772 MVV vector particles were harvested by ultracentrifugation and the viral pellet was fixed in equal volume of 5% glutaraldehyde supplemented with 0.2 M sodium cacodylate. The fixed viral pellet was embedded in 2% low melting point agarose. This process is designed to protect the integrity of the viral particles in the pellet, which was then fixed in 1% osmium tetroxide and stained with 1% uranyl acetate. The pellet was gradually dehydrated and finally embedded in 100% epoxy resin. The final steps of the protocol (sectioning and visualisation of the virions on an electron microscope) was carried out by Dr. Raffaella Carzaniga from the Electron Microscopy core facility at The Francis Crick Institute.

The electron micrographs of MVV WT-KV1772 virions showed even distribution of viral particles in all the sections analysed. Most MVV WT-KV1772 (Figure 4-13) particles were mature virions, with a fully processed capsid core encasing electron-dense material (presumably the viral RNA genome). The virions were sized between

100 nm and 150 nm in diameter. Different shapes were observed for the mature capsid core, including fullerene cone, cylindrical and a spherical form. No immature virions were observed in all the electron micrographs analysed. The virion morphology observed is in agreement with the previously described morphological features of MVV purified from infected sheep (Weiland and Bruns, 1980, Gendelman et al., 1986). This protocol worked remarkably well for WT-KV1772 particles, but several technical problems, including several sections of the resin containing only debris without intact viruses, were encountered with testing the mutants, which made it highly unreproducible. Therefore, this procedure will require further optimisation before it can be applied to study and compare IN mutants.

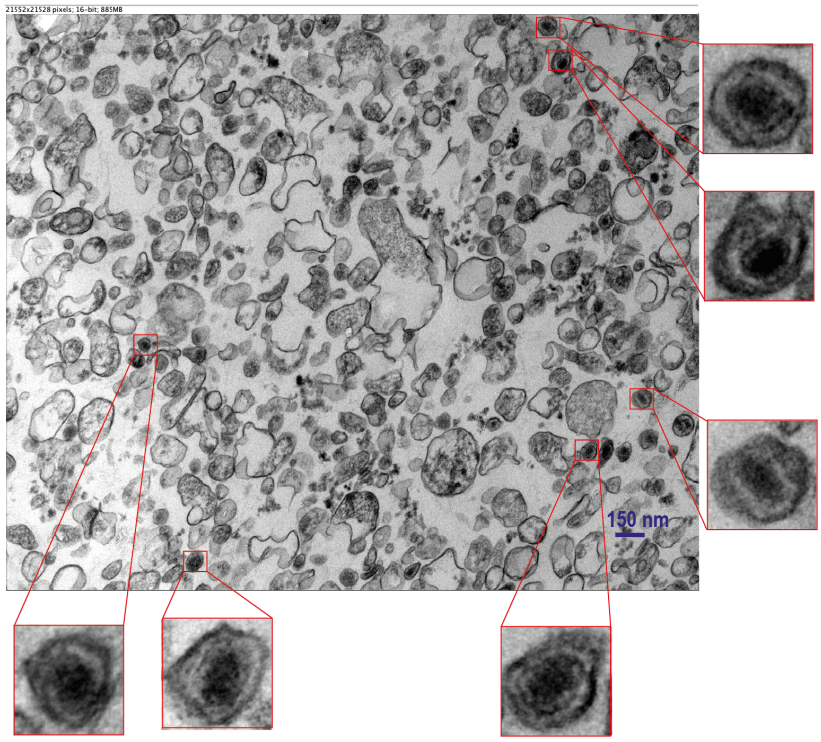

Figure 4-13 Visualisation of concentrated MVV viral particles.

Wild type KV-1772 MVV vectors were prepared for thin-section EM and a representative electron micrograph is shown. The coloured boxes show enlargements of individual virions. Electron micrographs were analysed using Fiji open-source platform (Schindelin et al., 2012, Hare et al., 2010).

4.6 Conclusions

Infectivity of the MVV IN mutants were tested in the context of the single-cycle vector system. Three mutations targeting the CTD-CTD interface interactions, F223A, Y261A and V263E and one mutation, QQQ211PGG, designed to disrupt integrity of the α-helical CCD-CTD linker, were introduced into the IN coding region of the modified MVV packaging construct. As expected, the control mutants E154Q and H12N were defective for integration, measured by the luciferase expression in the infected cells, 7 d post-infection. The mutants introduced in the CTD region as well as the CCD-CTD linker were largely defective for strand transfer activity *in vitro* (Figure 3-9). Their activities *in vitro* were measured using a pre-processed 19 bp donor DNA substrate with a sequence of the MVV U3 end of the LTR region, and the intasomes assembled with these mutant IN proteins formed with highly reduced stability in comparison to WT-KV1772 (Figure 3-12). DNA foot-printing performed with MLV and HIV-1 vDNA using Mu phage transposase, indicated a level of protection by an extended protein complex for 200-500 bp of vDNA, from both U3 and U5 ends (Wei et al., 1997, Chen et al., 1999). Thus, in principle, the intasome might be stabilised due to additional protein-DNA and protein-protein interaction. However, the MVV IN mutants generated in this study were greatly defective in the vector infectivity assay, behaving similar to the negative control mutants (Figure 4-9). The mutations introduced were outside the CIC, thus are not expected to directly interfere with the enzymatic activities of the intasome. The observations reported here strongly suggest these mutants were unable to assemble into intasomes, thereby unable to protect the vDNA and causing severe integration defects.

Although the observations made both *in vitro* (Figures 3-9 and 3-12) and in infected cells (Figure 4-9) indicate defects in integration caused solely by not being able to form intasomes, the potential pleiotropic defects caused by introducing mutations in the IN region were also tested. The mutants ability to perform reverse transcription was tested by quantifying the vDNA quantities at 2, 8 and 24 h post-infection. Two different methods of quantifying vDNA were employed. In one case, the early reverse transcription products were quantified, using primers annealing within *luc* gene. In the other case, the LRT product was detected by qPCR, using primers amplifying LTR-*gag* region (Figure 4-11A). The initial method to quantify vDNA by qPCR with

primers targeting the *luc* gene sequence of the proviral DNA yielded with potential non-specific amplification of residual plasmid DNA, still present in the viral suspensions used for infections, as well as incomplete inhibition of MVV reverse transcription by AZT (Figure 4-10).

The LRT strategy of quantifying vDNA enabled amplification of specific reverse transcription products, thereby evading the amplification of residual plasmid DNA. The LRT PCR quantification of vDNA 2 h, 8 h and 24 h post-infection did not reveal dramatic reverse transcription defects caused by the mutants (Figure 4-12). The general trend of vDNA synthesis and degradation was in accordance with previously observed HIV-1 data (Butler et al., 2002, Yoder et al., 2006). However, F223A, which had the reduced Gag-processing (Figure 4-2), also had an increased spike in vDNA synthesis at 24 h post-infection (Figure 4-12). Interestingly, the vDNA synthesis behaviour of H12N was not in accordance to the previously published data (Wu et al., 1999). The electron micrographs of WT-KV1772 virions were evenly distributed (Figure 4-13) and the morphological features of the mature WT-KV1772 virions were in accordance with previously published data. This protocol includes heavy metal staining of the viral particles and imaging 70-nm sections of the viral pellet embedded in a highly hydrophobic epoxy resin, with the critical step of finding the precise area of the pellet to section and image. This method fortuitously worked with the WT-KV1772 vectors and requires further optimisation to visualise the mutants.

The infectivity data presented in this section are consistent with the importance of the hexadecameric assembly of the MVV intasome for its integration activities and appear to validate the cryo-EM structure. The mutations introduced in the IN-coding region of the MVV *gag-pol* to target the CTD-CTD interface interactions and destabilise the CCD-CTD linker, led to severe infection defects while not abrogating reverse transcription. These observations do not address the functional significances of the extended intasome architecture, which will require a follow up study. Our working hypothesis is that the extended architecture may allow the intasome to form high-avidity/stoichiometry interactions with chromatin-bound factors, such as LEDGF/p75, potentially making lentiviruses more sensitive to local epigenetic environment. To begin to address this question, the following chapters 5 and 6 aim to understand the role of LEDGF/p75 in MVV integration.

Chapter 5. Testing the importance of LEDGF/p75 for MVV intasome assembly and viral infection

5.1 Aims

The host factor, LEDGF/p75, plays a critical role in targeting HIV-1 integration towards actively transcribing genes during viral infection (Ciuffi et al., 2005, Shun et al., 2007) and it was shown to interact with a range of lentiviral IN proteins (Cherepanov et al., 2003, Cherepanov, 2007). The proposed model of LEDGF/p75 binding to the MVV intasome involved sixteen IN CCDs interacting with sixteen LEDGF/p75 IBDs (Supplementary figure 14, Ballandras-Colas et al., 2017). Recent results from our laboratory confirmed that MVV intasome can recruit sixteen LEDGF/p75 subunits (Ballandras-Colas et al., Manuscript in preparation). Similar to LEDGF/p75, HRP2 was shown to contain both the PWWP and IBD domains (Cherepanov et al., 2004, Vanegas et al., 2005). Both LEDGF/p75 and HRP2 were implicated in playing a role in repairing DNA double-strand breaks (DSBs) by performing homologous end joining (Daugaard et al., 2012, Baude et al., 2016). In the absence of LEDGF/p75, HRP2 was reported to aid in targeting HIV-1 integration towards actively transcribing genes (Schrijvers et al., 2012a, Schrijvers et al., 2012b, Wang et al., 2012), elaborated in section 1.6.2.

The aim for this part of the project was to test the importance of LEDGF/p75 for MVV intasome assembly and infectivity. In order to test the requirement for LEDGF/p75 for MVV infectivity, single-cycle vector infections were carried out in HEK-293T cells that were knocked out for both LEDGF/p75 and HRP2. Infectivity of MVV WT and the IN active site mutant form, E154Q was monitored by measuring luciferase expression.

5.2 Testing the importance of LEDGF/p75 for MVV integration

In agreement with previously published data (Ballandras-Colas et al., 2017), no FS integration products were observed in the *in vitro* IN strand transfer activity without LEDGF/p75 in the reaction mixture (Figures 3-8, 3-9, and Figure 5-1).

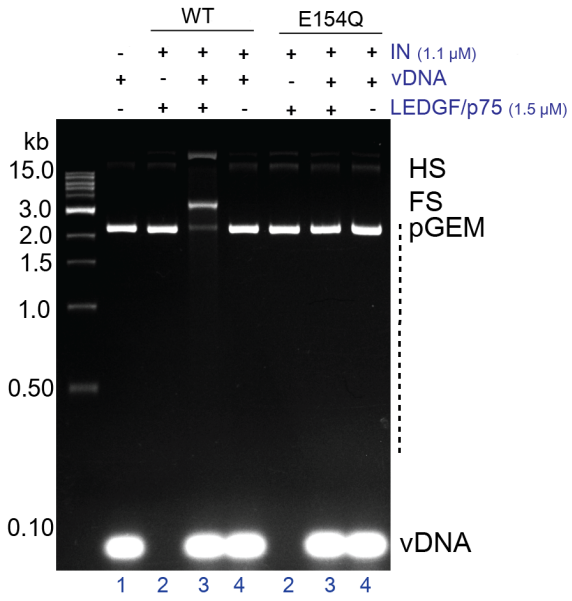

Figure 5-1 LEDGF/p75 is critical for MVV IN strand transfer activity in vitro.

Strand transfer activities of WT and E154Q MVV IN at a concentration of 1.1 μM, incubated with 1.5 μM LEDGF/p75 or buffer, with or without 20 μM vDNA, and in the presence of 7.5 ng/μl target DNA. Deproteinized reaction products were separated in 1.5% agarose gels and detected by staining with ethidium bromide. Migration positions of the reaction products, vDNA and pGEM are indicated to the right of the gel; positions of DNA molecular size markers (kb) are shown on the left. Lane 1 is a mock reaction that does not contain IN and LEDGF/p75, lanes 2 do not contain vDNA, lanes 3 contain all the components and lanes 4 do not contain LEDGF/p75.

5.2.1 Analysis of HEK-293T LKO and LHKO cells

Parental HEK-293T cells and cells deficient for LEDGF/p75 and HRP2 expression through genetic knockout (KO), henceforth referred as LKO (LEDGF/p75 KO), HKO (HRP2 KO) and LHKO (double LEDGF/p75 and HRP2 KO), were obtained from Prof. Eric M Poeschla, University of Colorado Denver. The construction of the LKO cell line, using transcription activator-like effector nucleases (TALENs) was described in (Fadel et al., 2014). The cell lines HKO and LHKO are unpublished. To confirm the phenotypes, the cells were analysed by Western blotting for the expression of LEDGF/p75, HRP2 and housekeeping gene, β-actin. As expected, LKO and HKO cells lacked detectable expression of LEDGF/p75 and HRP2, respectively. LHKO cells did not express LEDGF/p75 or HRP2 (Figure 5-2).

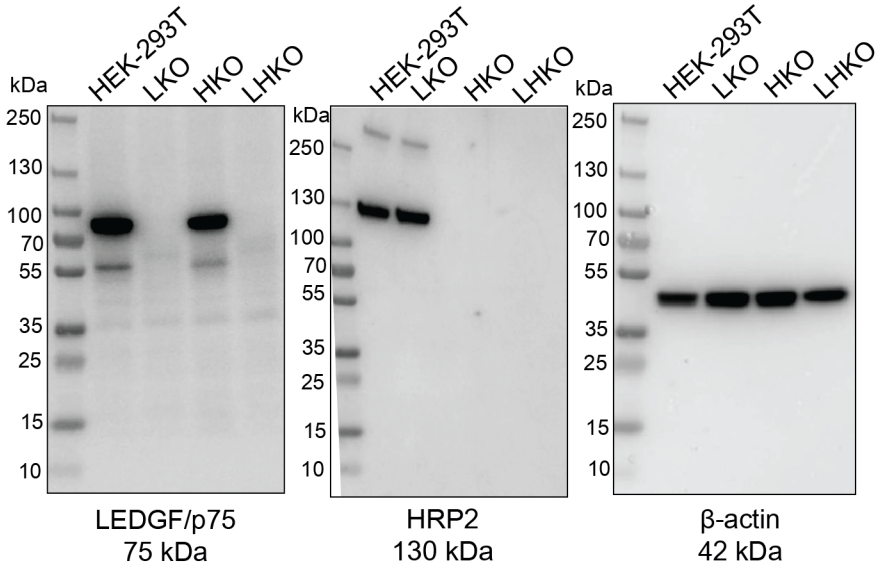

Figure 5-2 Testing the expression of LEDGF/p75 and HRP2 in LKO and LHKO cells.

Cells were tested for the presence of LEDGF/p75 and HRP2. Total cell extracts of HEK-293T, LEDGF/p75-null (LKO), HRP2-null (HKO), and LEDGF/p75- and HRP2-null (LHKO) cells were separated by SDS-PAGE and tested by Western blotting with anti-LEDGF/p75, anti-HRP2, and anti-β-actin antibodies. The molecular masses of the proteins tested are indicated below the blots.

5.2.2 MVV infectivity in LKO and LHKO cells

MVV WT and E154Q vectors were produced by transfection of HEK-293T cells and quantified by Taq-PERT (sections 2.4.1 and 2.6.1). HEK-293T, LKO and LHKO cells were seeded in 48-well plates and infected with 0.5 mU RT-normalised virus. 2 d post-infection, cells were expanded into 6-well plates and harvested 7 d post-infection. Cell extracts were prepared as described in section 2.6.3 and luciferase activity was measured.

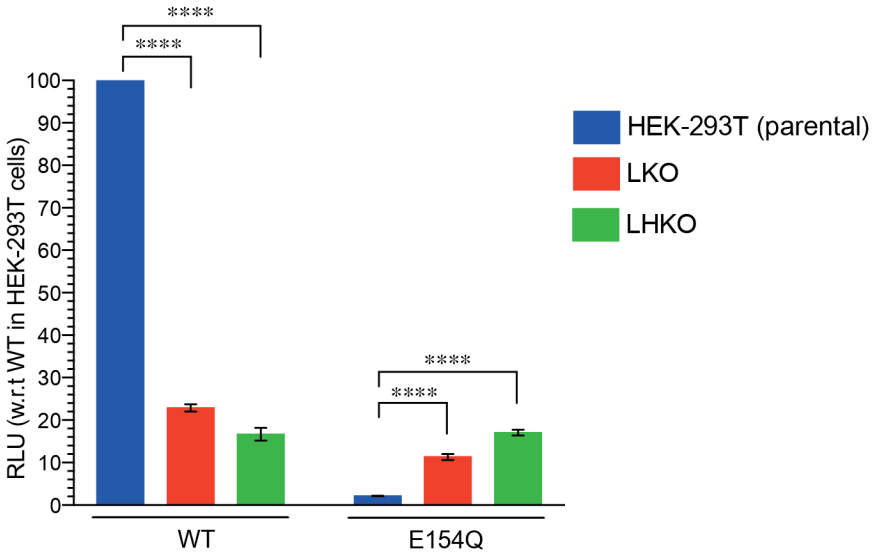

Figure 5-3 Infection in LKO and LHKO cells with MVV vectors.

HEK-293T, LKO and LHKO cells were infected with 0.5 mU RT-normalised MVV WT and E154Q vectors. Luciferase activity was measured 7 d post-infection, with RLU normalised to WT infectivity in HEK-293T cells. Means with standard error of the mean plotted of three biological repeats, with two technical repeats each; Two-tailed paired t-tests (individual comparisons), $p<0.0001$ (****).

Infectivity levels of the viruses were normalised with respect to WT in HEK-293T cells, which was set to 100%. Similar to the observations made in HIV-1 infection (Wang et al., 2012), WT infectivity reduced by half in LKO cells and by a further 20% in LHKO cells. E154Q was defective for integration in HEK-293T cells (Figure 5-3), in

agreement with observations made in Chapter 4 (Figures 4-8 and 4-9). Surprisingly, the luciferase expression of E154Q infected LKO and LHKO cells significantly increased to 12% and 18%, respectively, compared to the infectivity of E154Q in HEK-293T cells (Figure 5-3). This increase in luciferase expression hinted at an alternative path of vDNA introduction into the host genome and required additional validation.

5.2.3 MVV infectivity in the presence of Dolutegravir (DTG)

INSTIs prevent the catalytic step of strand transfer by competing with target DNA to bind to the intasome active sites (Engelman, 2019), while IN active site mutants are unable to perform the catalytic steps of the integration reaction. To validate the increase in infectivity of the integration incompetent virus, E154Q virus (Figure 5-3), infections were carried out using WT vector in the presence of varying concentration of DTG. It is an IN strand transfer inhibitor (INSTI) described in section 1.9.1, that was shown to inhibit MVV IN strand transfer activity *in vitro* (Ballandras-Colas et al., 2017).

HEK-293T, LKO and LHKO cells were seeded in 48-well plates and infected with 0.5 mU RT-normalised virus, in the presence of 10,000, 1,000, 100, 10, 1 nM DTG. The INSTI stock was prepared in 100% DMSO. Therefore, a control (DTG-free) infection was carried out in the presence of equivalent concentration of DMSO (0.1% final). Infected cells were treated as described in section 2.6.6 and luminescence was measured 7 d post-infection.

As expected, DTG strongly suppressed infectivity of WT MVV vector. WT infectivity in HEK-293T cells decreased from 100% to ~3% with increasing amounts of DTG, while this effect was less pronounced in LKO and LHKO cells, with infectivity reducing from 30 to 50% in DMSO treated cells, to ~10% in 10,000 nM DTG (Figure 5-4A). WT infectivity in HEK-293T cells reduced significantly with every 10-fold increase in the amount of DTG, with a dramatic 5-fold decrease in infectivity from 100 nM DTG to 1,000 nM DTG (Figure 5-4A). Reduction in WT infectivity in LKO cells with increasing amounts of DTG was not significantly different between 10 nM to 100 nM and 1,000 nM to 10,000 nM DTG. WT infectivity in LHKO cells, in most cases was not significantly different with increasing amount of DTG (Figure 5-4A).

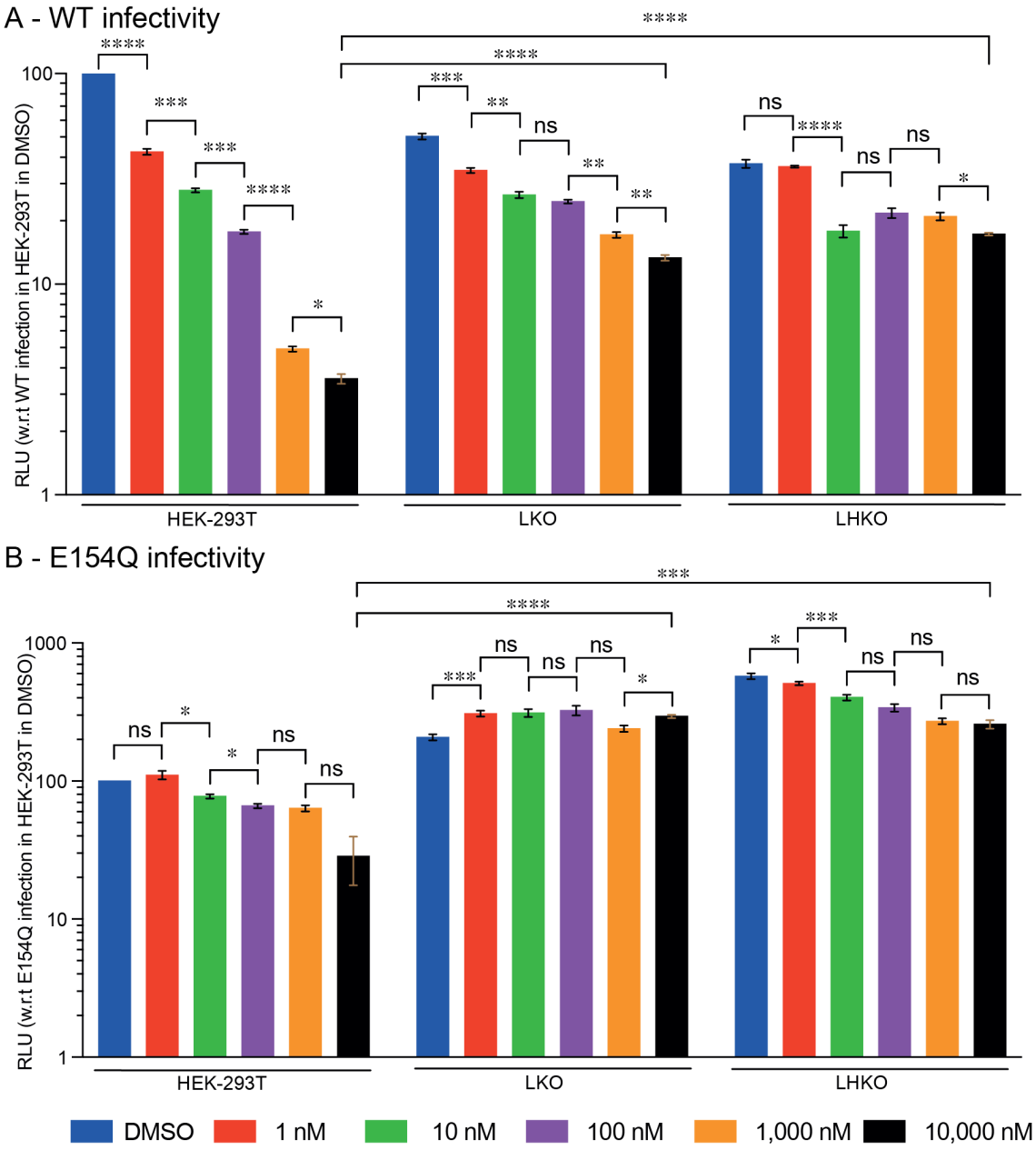

Figure 5-4 Effect of DTG on MVV infectivity.

HEK-293T, LKO and LHKO cells were infected with 0.5 mU RT-normalised MVV WT-KV1772 and E154Q viral vectors in the presence of DMSO or 1, 10, 100, 1,000, 10,000 nM DTG, an INSTI. Luciferase measured 7 d post-infection in (A) WT and (B) E154Q infected cells. Means with standard error of the mean plotted of three biological repeats, with two technical repeats each; Two-tailed paired t-tests (individual comparisons). p=0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****).

WT infectivity in the presence of 10,000 nM DTG (~3% RLU) was considered as complete inhibition of IN activity, since it is comparable to E154Q infectivity in HEK-293T cells in the absence of drug (Figures 5-3 and 5-4). Even at the highest concentration of the drug tested, 10,000 nM DTG, WT infectivity in LKO and LHKO cells was significantly higher than in HEK-293T cells (Figure 5-4A). This observation was in contrast to the infectivity in these cells in the absence of DTG (Figure 5-3).

Given that E154Q, a substitution within the essential catalytic D,D-35,E triad, completely disrupts IN catalytic activities, no additional effect of DTG on the infectivity of E154Q virus is expected. Furthermore, INSTIs engage the pair of catalytic Mg^{2+} ions in the IN active site; because E154Q disrupts Mg^{2+} ion coordination, the active site IN mutant is not expected to bind DTG. Concordantly, luciferase expression in cells infected with E154Q were generally unaffected by the presence of DTG, with an increased luciferase expression in LKO and LHKO cells, as seen in Figure 5-3. Even in the presence of 10,000 nM DTG, the luciferase expression in LKO and LHKO cells was significantly higher than luciferase expression in HEK-293T cells (Figure 5-4B).

The increase in luciferase expression of integration defective MVV, either the IN active site mutant E154Q or WT inhibited by DTG observed in LKO and LHKO cells is intriguing. One attractive explanation is that LEDGF/p75 and possibly HRP2 are important for the stability of the intasome. In the absence of these host factors, vDNA ends may become highly susceptible for non-homologous end joining machinery (NHEJ), which inserts them into chromosomal DNA in an IN-independent fashion.

5.3 Proviral DNA quantification

Lentiviral integration mediated by the viral IN is targeted towards transcriptionally active chromosomal locations (Schröder et al., 2002). Measuring luciferase expression that requires transcription of *luc* gene maybe restricted to vDNA insertion events occurring in euchromatin alone. Ku80 is part of the NHEJ DNA repair machinery (Yang et al., 2016), which was shown to be recruited to repair DSBs in heterochromatin (Tsouroula et al., 2016). The vDNA insertion events occurring in the heterochromatin via NHEJ may not be detected by measuring luciferase expression

alone. In order to measure the sum total of vDNA insertion in infected cells, proviral DNA was quantified.

5.3.1 Quantification of *luc* gene to measure proviral DNA

The *luc* gene is not endogenously present in parental, LKO or LHKO cells, with only infected cells expressing luciferase. As a measure of viral copy number, primers targeting the *luc* gene were used in a TaqMan probe based qPCR quantification. HEK-293T, LKO and LHKO cells were infected with WT and E154Q MVV. The infected cells were harvested 7 d post-infection. As a negative control, infections were also carried out in the presence of 100 μM AZT, an RT inhibitor, described in section 1.9. DNA was purified from infected cells and the relative viral copy number was quantified from a standard curve generated using pCVW-CG-Luc plasmid, elaborated in section 2.6.4.

Unexpectedly, HEK-293T, LKO and LHKO cells infected with WT did not display statistically significant differences in the amount of proviral DNA present 7 d post-infection (Figure 5-5), unlike the differences observed in luciferase expression (Figures 5-3 and 5-4A). Comparing the results obtained in the presence of AZT pointed out to a major technical problem. The amount of proviral DNA present in cells infected with WT in the presence of AZT was comparable to the amount seen in its absence (Figure 5-5). This strongly indicated that, despite extensive DNase treatment, the MVV vectors retained considerable amounts of transfected plasmid DNA, which also contain the *luc* gene. Compared to HIV-1 derived vectors, the MVV system produces relatively low vector titres. As a consequence, the virus must be concentrated, which exacerbates the contamination with input viral DNA constructs.

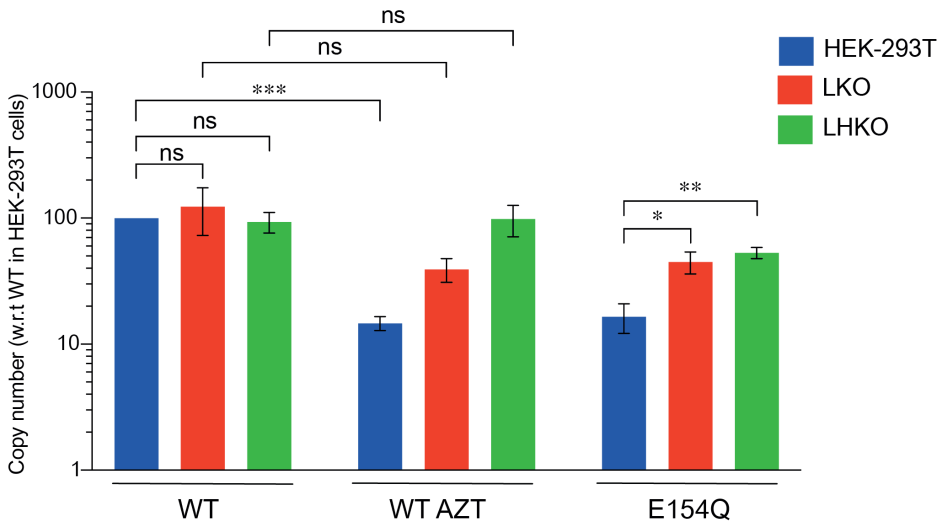

Figure 5-5 Quantification of proviral DNA using *luc* real-time qPCR.

HEK-293T, LKO and LHKO cells were infected with 0.5 mU RT-normalised MVV WT and E154Q viral vectors. DNA was purified 7 d post-infection and the amount of vDNA was quantified by qPCR, with primers annealing within the *luc* gene. Infection of WT in the presence of 100 μM AZT (nucleoside RT inhibitor) was used as a negative control. Quantities of the viral cDNA were determined from a standard curve generated using pCVW-CG-Luc. Values were normalised to WT vDNA synthesis in HEK-293T cells, which was set to 100%. Means plotted with standard error of the mean from two biological replicates, with two technical repeats each; Two-tailed paired t-tests (individual comparisons). p=0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****).

5.3.2 Quantification of proviral DNA by LRT PCR

To specifically quantify vDNA produced through reverse transcription from the background of variable contamination by the transfected plasmid, the LRT method was used, elaborated in section 4.5.1.2. HEK-293T, LKO and LHKO cells were infected with DNase treated MVV WT and E154Q mutant viruses at a titre of 2.5 mU RT. As a negative control, HEK-293T cells were infected with WT virus in the media supplemented with 100 µM ddC, a nucleoside RT inhibitor, shown to be highly active against MVV (Thormar et al., 1993). Infected cells were harvested 8 h post-infection and 7 d post-infection and DNA was extracted. A TaqMan probe based qPCR was performed with primers targeting the MVV U3 region and 5′ region of *gag* (Section 2.6.5), described in section 4.5.1.2.

The proviral DNA in HEK-293T cells infected with WT was normalised as 100% at both the tested time points post-infection. As expected, addition of ddC suppressed vDNA synthesis. At the earlier time point of 8 h post-infection (Figure 5-6), the vDNA quantities in HEK-293T cells infected with WT and E154Q were similar to the observations made in Figure 4-12. An increase in vDNA quantities was observed in LKO and LHKO cells infected with either WT and E154Q (Figure 5-6).

At the later time point of 7 d post-infection, the proviral DNA reduced in LKO and LHKO cells infected with WT (Figure 5-6). An increase in the amount of proviral DNA was observed in LKO and LHKO cells infected with E154Q, which is consistent with increased luciferase signals. Quantification of proviral DNA required further optimisation, with additional infection repeats to determine statistically significant values.

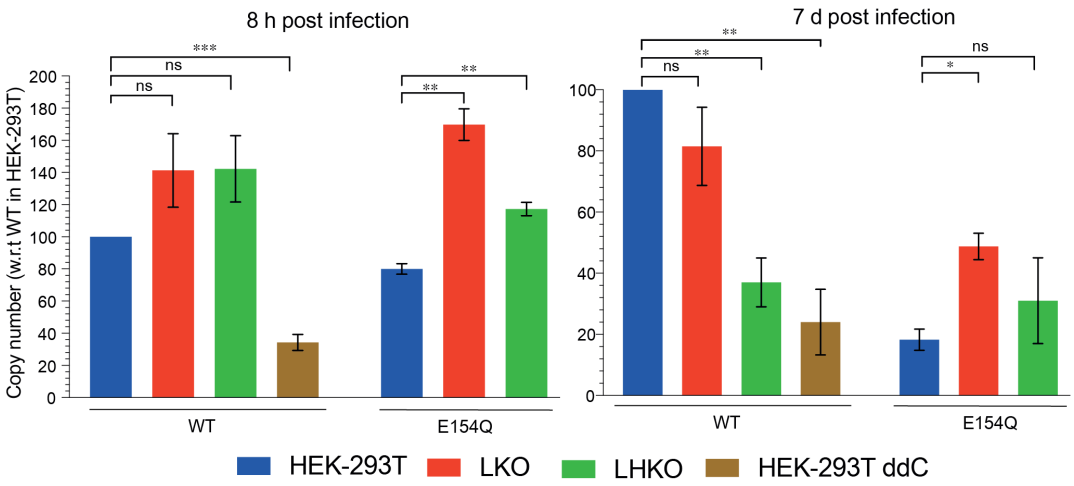

Figure 5-6 Proviral DNA quantification by LRT real-time qPCR.

HEK-293T, LKO and LHKO cells were infected with 2.5 mU RT-normalised MVV WT-KV1772 and E154Q viral vectors. DNA was purified 8 h and 7 d post-infection and the amount of vDNA was quantified by qPCR, with primers annealing within MVV U3 and *gag*. Infection with WT vector in the presence of 100 µM ddC (nucleoside RT inhibitor) served as a negative control. Quantities of the vDNA were determined from a standard curve generated using pCR4-TOPO-MVV-LRT plasmid. Data are normalised to WT vDNA synthesis in HEK-293T cells 8 h and 7 d post-infection which were set to 100%. Means plotted with standard error of the mean from four biological replicates; Two-tailed paired t-tests (individual comparisons). p=0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****).

5.4 Re-expression *Oa*LEDGF/p75 in HEK-293T LHKO cells

The MVV infection phenotype observed in the absence of LEDGF/p75 and HRP2 in HEK-293T cells indicated that the host factors might play an important role in intasome assembly and stability. In such case, it should be possible to reverse the phenotype upon re-introduction of at least one of these host factors in LHKO cells.

5.4.1 Sequences of Sheep and Human LEDGF/p75

Amino acid sequences of *Homo sapiens* (human) and *Ovis aries* (sheep) LEDGF/p75 proteins were 97% identical, with their IBDs being 100% identical (Appendix figure 8.4) and sheep is the host organism of MVV. Two forms of *Oa*LEDGF/p75, WT and D366N, were introduced into LHKO cells, henceforth referred as LHKO$_{(Oa\text{LEDGF_WT})}$ and LHKO$_{(Oa\text{LEDGF_D366N})}$. LEDGF/p75 D366N is an IN binding mutant, which does not interact with lentiviral INs (Cherepanov et al., 2005b, Cherepanov, 2007).

5.4.2 Generation of LHKO$_{(Oa\text{LEDGF_WT})}$ and LHKO$_{(Oa\text{LEDGF_D366N})}$ cells

DNA fragments encoding WT and D366N *Oa*LEDGF/p75 were ligated into pQCXIP retroviral vector, by Prof. Peter Cherepanov. HEK-293T cells were co-transfected with MLV pCG-Gag-Pol packaging construct, pCG-VSV-G (Ulm et al., 2007) and pQFLAG-Puro-LEDGF/p75-WT or D366N. The resulting retroviral vectors were used to transduce LHKO cells. Transduced cells were selected for *Oa*LEDGF/p75 expression in the presence of 1 µg/ml puromycin. *Oa*LEDGF/p75 expression in the cells was tested by Western blotting (Figure 5-7), where presence of either WT or D366N forms of *Oa*LEDGF/p75 were detected in LHKO cells.

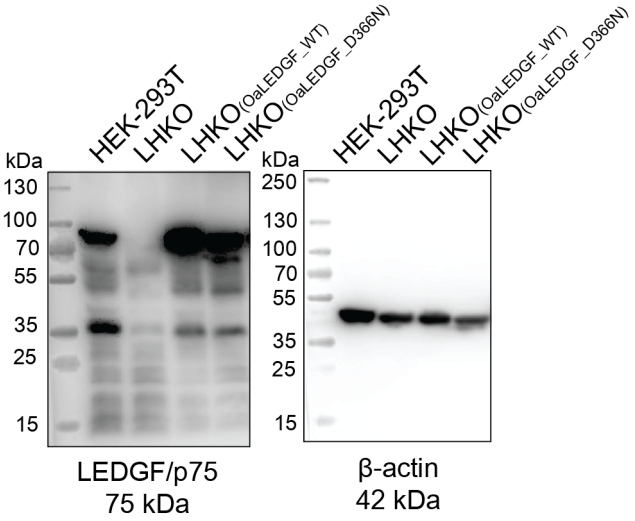

Figure 5-7 Stable ectopic expression of OaLEDGF/p75 in LHKO cells.

Whole cell extracts of LHKO cells stably transduced with retroviral vectors expressing WT or D366N forms of ovine LEDGF/p75 (*Oa*LEDGF/p75) were tested for expression of the proteins by Western blotting. 10 µg (total protein) each of HEK-293T, LHKO, LHKO$_{(OaLEDGF_WT)}$ and LHKO$_{(OaLEDGF_D366N)}$ cell lysate was loaded per well and tested with anti-LEDGF/p75 or anti-β-actin antibodies. The molecular masses of the proteins tested are indicated below the blots.

5.4.3 MVV infectivity in *Oa*LEDGF/p75 expressing LHKO cells

HEK-293T, LKO, LHKO, LHKO$_{(Oa\text{LEDGF_WT})}$ and LHKO$_{(Oa\text{LEDGF_D366N})}$ cells were infected with 0.5 mU RT-normalised WT or E154Q viruses. Infected cells were harvested 7 d post-infection and expression of luciferase was measured, as described in section 2.6.3. In agreement with earlier observations (Figures 5-3 and 5-4), MVV infectivity in LKO and LHKO cells reduced significantly, compared to the parental HEK-293T cell line, while luciferase expression of LKO and LHKO cells infected with E154Q increased (Figure 5-8).

WT infectivity in LHKO$_{(Oa\text{LEDGF_WT})}$ cells was only partially rescued, with a 10% increase in luciferase expression in comparison to its infectivity in parental LHKO cells, and WT infectivity in LHKO$_{(Oa\text{LEDGF_D366N})}$ cells reduced by 20% (Figure 5-8). The reason for the decrease in infectivity of LHKO$_{(Oa\text{LEDGF_D366N})}$ cells was unclear,

since the D366N form of LEDGF/p75 does not interact with IN. It is important to note that neither of these differences was statistically significant. However, in cells infected with E154Q, the RLU of LHKO$_{(OaLEDGF_WT)}$ significantly decreased, potentially confirming the role of LEDGF/p75 in suppressing infectivity of integration deficient virus (Figure 5-8). Conversely, destabilisation of the intasome by abrogation of LEDGF/p75 could lead to increased availability of vDNA ends for integration in an IN-independent manner, potentially by the host DNA repair machinery NHEJ.

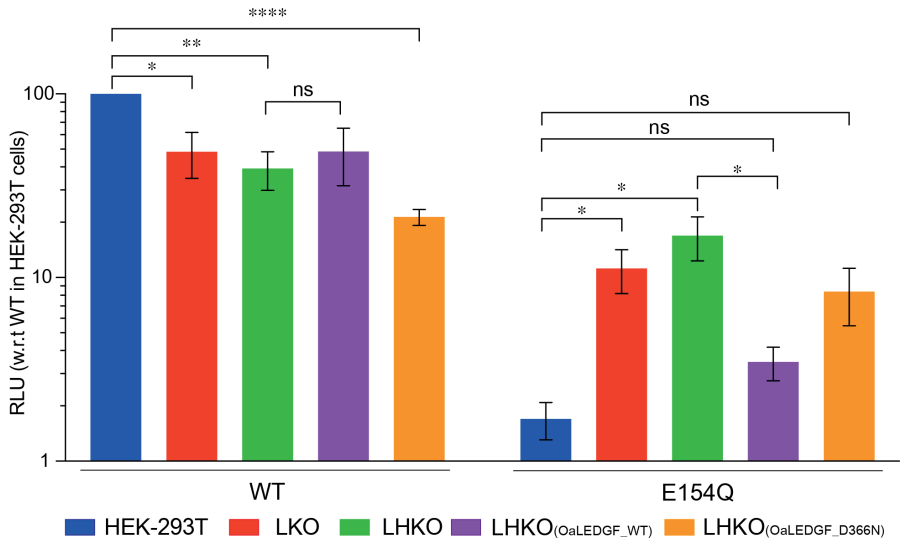

Figure 5-8 MVV vector infectivity in cells expressing OaLEDGF/p75.

HEK-293T, LKO, LHKO, LHKO$_{(OaLEDGF_WT)}$ and LHKO$_{(OaLEDGF_D366N)}$ cells were infected with 0.5 mU RT-normalised MVV WT and E154Q viral vectors. Luciferase activity was measured 7 d post-infection, with RLU normalised to WT infectivity in HEK-293T cells. Means with standard error of the mean plotted of three biological repeats, with two technical repeats each; Two-tailed paired t-tests (individual comparisons). p=0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****).

5.5 MVV integration sites in HEK-293T, LKO and LHKO cells

Lentiviruses display strong preference for integration within bodies of actively transcribed genes, while disfavouring vicinities of transcription start sites (TSSs) and CpG islands (Schröder et al., 2002, Ballandras-Colas et al., 2017). In case of HIV-1, these traits were shown to depend on the interaction between IN and LEDGF/75, and to a lesser extent HRP2. In addition, the interactions between the intasome and target DNA result in weak local base preferences, which become apparent when many integration sites are aligned (Holman and Coffin, 2005, Ballandras-Colas et al., 2017). A shift in HIV-1 integration away from transcription units was observed in the absence of LEDGF/p75, and the effect enhanced under conditions of HRP2 depletion. However, the local DNA sequence preference was not impacted (Shun et al., 2007, Schrijvers et al., 2012b), elaborated in section 1.6.2.

The data reported in this chapter showed that MVV retained considerable integration activity in the absence of LEDGF/p75 and HRP2. To determine if the host factors impact MVV integration preferences, MVV integration sites in parental, LKO and LHKO cells were amplified and sequenced.

5.5.1 Infections and preparation of genomic DNA samples for integration site sequencing

HEK-293T, LKO and LHKO cells were infected with fresh MVV WT viral preparations. Infected cells were expanded into 10-cm dishes and harvested 5 d post-infection. GFP expression of infected cells was determined by flow cytometry. Live cells were gated from the total cell population, where single cells were analysed for GFP expression. The gating strategy was based on non-infected, GFP-negative HEK-293T cells and the number of GFP-positive cells were plotted in a bar graph (Figure 5-9). DNA was purified from the infected cells and samples were processed further by Dr. Parmit Singh (Alan Engelman lab, Dana Farber Institute, Boston, MA, USA).

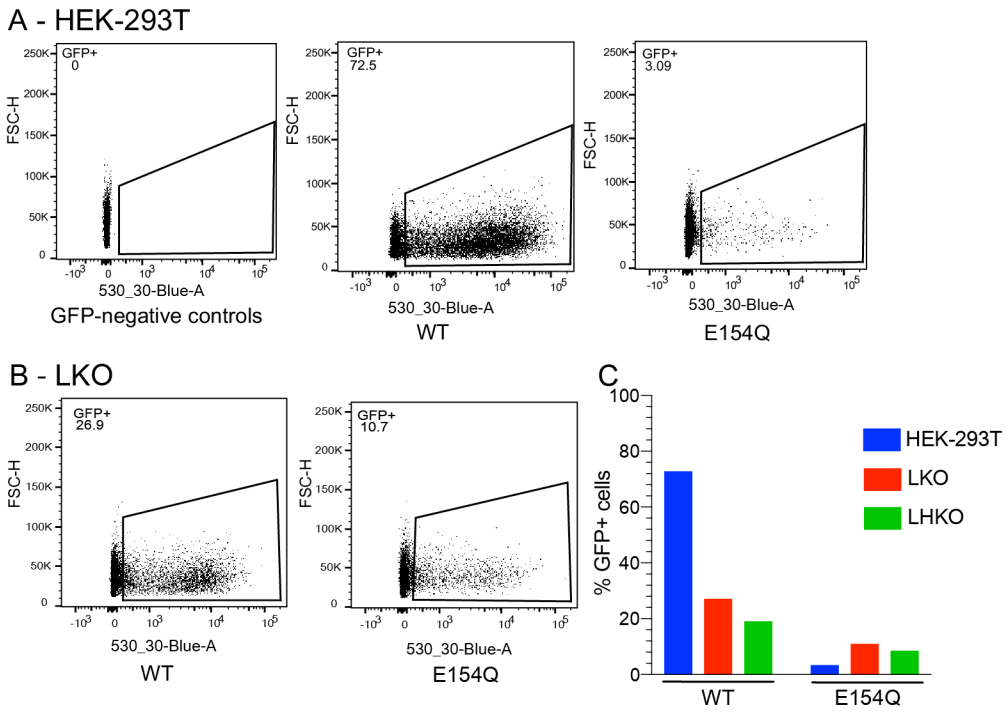

Figure 5-9 Infection of HEK-293T, LKO and LHKO cells with MVV vectors for integration site sequencing analysis.

HEK-293T, LKO and LHKO cells were infected with fresh preparations of concentrated MVV WT and E154Q viral vectors and harvested 4 d post-infection. The gating strategy used for infected (A) HEK-293T cells and (B) LKO cells. Data for infected LHKO cells is not shown. (C) Bar graph plotted with the percentage of GFP-positive cells, calculated from analysing 10,000 single cells.

The genomic DNA was digested with an endonuclease, *Mse*I (5′-TTAA-3′), which generates a 5′-TA overhang. The MVV LTR is 418 bp long, with an *Mse*I site 215 bp upstream of the U5 end. The digested genomic fragments were annealed to an asymmetric linker and DNA was amplified in a two-round nested PCR using LTR specific primers and the same linker specific primer with an adaptor DNA sequence at its 3′ end (Serrao et al., 2016) (Figure 5-10). The samples were submitted to Genewiz Laboratories for Illumina sequencing. The sequencing files were processed by Dr. Parmit Singh (Engelman lab) and Prof. Peter Cherepanov, (described in Serrao et al., 2016) and integration site maps were generated.

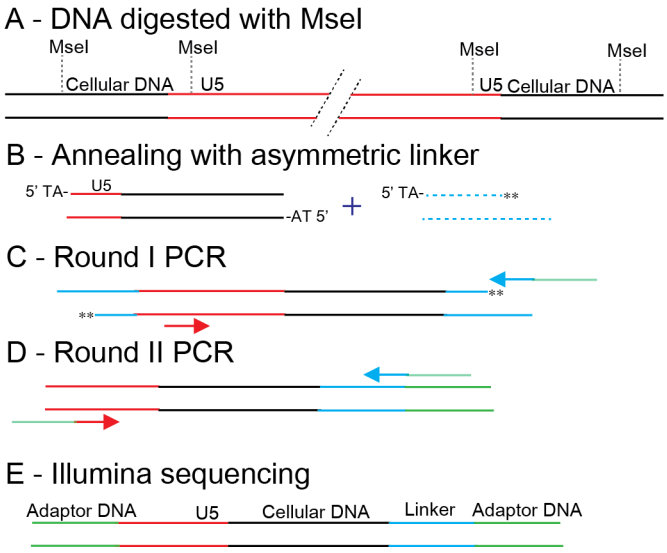

Figure 5-10 Illustration of the steps involved in sequencing MVV genomic integration sites.

(A) DNA purified from infected cells is digested with *Mse*I. (B) Digested DNA is ligated to an asymmetric linker. (C and D) Two rounds of nesting PCR. (E) PCR products analysed by Illumina sequencing (Adapted from Vanegas et al., 2005, Serrao et al., 2016).

5.5.2 Distributions of MVV integration sites in HEK-293T, LKO and LHKO cells

MVV WT integration in HEK-293T cells was generally targeted towards gene bodies (67%), ~1.5-fold higher than in the control MRC and away from TSSs (2.9%) and CpG islands (2.6%) (Table 5-1). A shift in integration pattern was observed in LKO and LHKO cells, wherein the integration in gene bodies was 47.7% and 49.1%, respectively, which were closer to the MRC values and ~1.5-fold lower than the observations made in HEK-293T cells (Table 5-1). A dramatic increase of integration events occurring in the vicinity of TSSs and CpG islands was also observed in LKO and LHKO cells, in comparison with HEK-293T cells (Table 5-1). The preference for MVV integration into heterochromatic lamina associated domains (LADs) and transcriptionally active speckle associated domains (SPADs) was in stark contrast to preference for these regions by HIV-1. MVV preference of integration into LADs was

similar in all the cell lines tested, and nearly equal to the MRC values (24.5%), with a 2-fold increase in integration into SPAD regions in the KO cells (Table 5-1). HIV-1 preference for integration into LAD and SPAD regions is driven by the CA-CPSF6 interaction and is a specific feature of primate lentiviruses (Li et al., 2020).

Table 5-1 HIV-1 and MVV integration site distributions in HEK-293T cells.

Virus or control[a]	Cell type	Unique sites	RefSeq genes (%)[b]	±2.5 kb TSS (%)[c]	±2.5 kb CpG (%)[c]	LADs (%)[b]	SPADs (%)[b]	Gene density[d]
MRC	*in silico*	265,399	45	3.8	3.8	24.5	3.7	8.5
HIV-1[e]	HEK-293T	43,230	83.3	4.2	5.5	5.7	29.8	20.9
	LKO	28,177	63.1	10.2	11.6	10.7	15.3	14.4
MVV	HEK-293T	411,721	67	2.9	2.6	19.2	3.8	9.1
	LKO	2,235	47.7	7.5	7.6	23.3	8	10.3
	LHKO	24,321	49.1	7.8	8.6	23.9	7.6	9.9

[a] Matched random control (MRC) or a single-cycle GFP vector (HIV-1 or MVV).

[b] Integration events inside a genomic feature (RefSeq gene, LAD, or SPAD).

[c] Integration events within 2.5 kb of a feature (TSS or CpG island).

[d] Average number of RefSeq genes within a 1-Mbp window (±500 kbp) of integration site.

[e] Integration sites from Sowd et al., 2016 and Li et al., 2020

HIV-1 had a strong preference for integration into gene dense regions, which reduced significantly in the absence of LEDGF/p75. MVV had no such preference for gene dense regions (9.1%), which remained consistent in LKO and LHKO cells, 10.3% and 9.9%, respectively (Table 5-1). Integration frequencies based on the expression level of the genes was also mapped. Gene expression in HEK-293T cells, analysed by RNAseq and previously described (Sowd et al., 2016, Li et al., 2020), was divided into 5 sets, with increasing levels of expression from 1 to 5. HIV-1 strongly prefers integration into highly expressing genes, that is mediated by the presence of LEDGF/p75 to a large extent. In the absence of LEDGF/p75, HIV-1 integration preference was still strongly towards highly expressed genes, potentially due to the presence of HRP2 and other host factor interactions (Figure 5-11A). On the other hand, MVV integration in HEK-293T cells into highly expressing genes was similar to the HIV-1 preference in LKO cells, which reduced further closer to the MRC values in LKO and LHKO cells (Figure 5-11A).

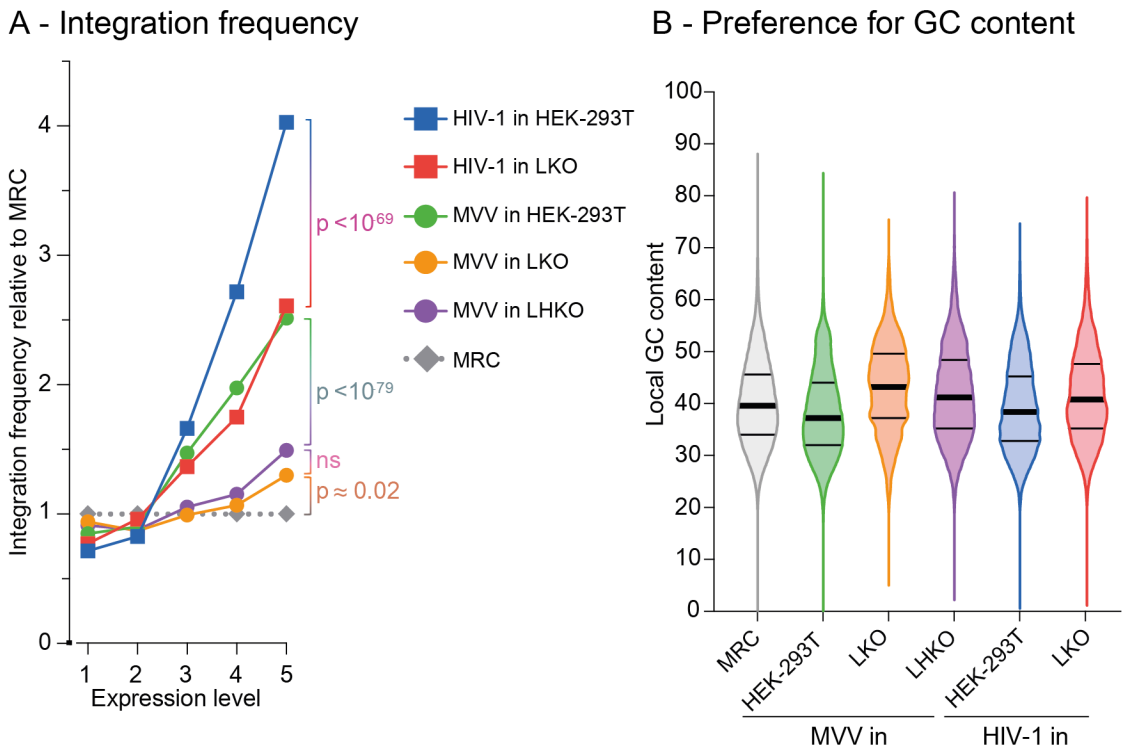

Figure 5-11 Transcription activity and local GC content at HIV-1 and MVV integration sites in HEK-293T, LKO and LHKO cells.

(A) Human genes were divided between 5 bins based on their expression level in HEK-293T cells (bins 1 and 5 contained genes with the lowest and highest expression, respectively). HIV-1 and MVV integration sites were scored in each bin. Note the strong preference of HIV-1 and MVV to integrate into more active TUs. By contrast, in silico-generated matched random control (MRC) integration sites shows equiprobable distribution between bins. (B) Genomic GC content within close proximity of integration sites (within 500 bp windows) is shown in form of a violin plot for each sample. Thick black lines indicate median values, and thin black lines demarcate quartiles. HIV-1 integration sites were from (Li et al., 2020, Hare et al., 2009a).

A nearly 5% increase in integration events targeted towards GC-rich regions was observed in LKO and LHKO cells infected with MVV, in comparison to HEK-293T cells. LEDGF/p75 is known to interact with the genomic DNA through its PWWP domain in particular, along with an NLS and two AT-hook motifs, (Turlure et al., 2006) elaborated in section 1.6.2. The AT-hooks are DNA binding protein motifs, that have an increased preference for A/T rich regions in the genome and in the absence of LEDGF/p75, it is unsurprising to observe an increase in integration targeted towards GC-rich regions of the chromatin. This LEDGF/p75 dependent increase in integration targeting towards GC-rich regions was also observed in HIV-1 (Figure 5-11B). The nucleotide sequence preferences of MVV in HEK-293T cells were mapped. As described in section 5.5.1, integration events were mapped from U5 vDNA ends alone. Based on the two-fold symmetry observed in the target DNA (positions -2 and -1, and 6 and 7) with a 6 bp gap from the site of U5 integration, the U3 vDNA end insertion was predicted (Figure 5-12). This integration site preference with the two-fold symmetry arises only as a direct consequence of the two-fold symmetry of the intasome complex (Ballandras-Colas et al., 2017). Interestingly, the palindromic nucleotide preference is reduced in LKO and LHKO cells, which could be a direct consequence of the assembly of a less stable intasome complex and this reduction is not observed in LEDGF/p75 and HRP2 depleted cells infected with HIV-1 (Shun et al., 2007, Wang et al., 2012).

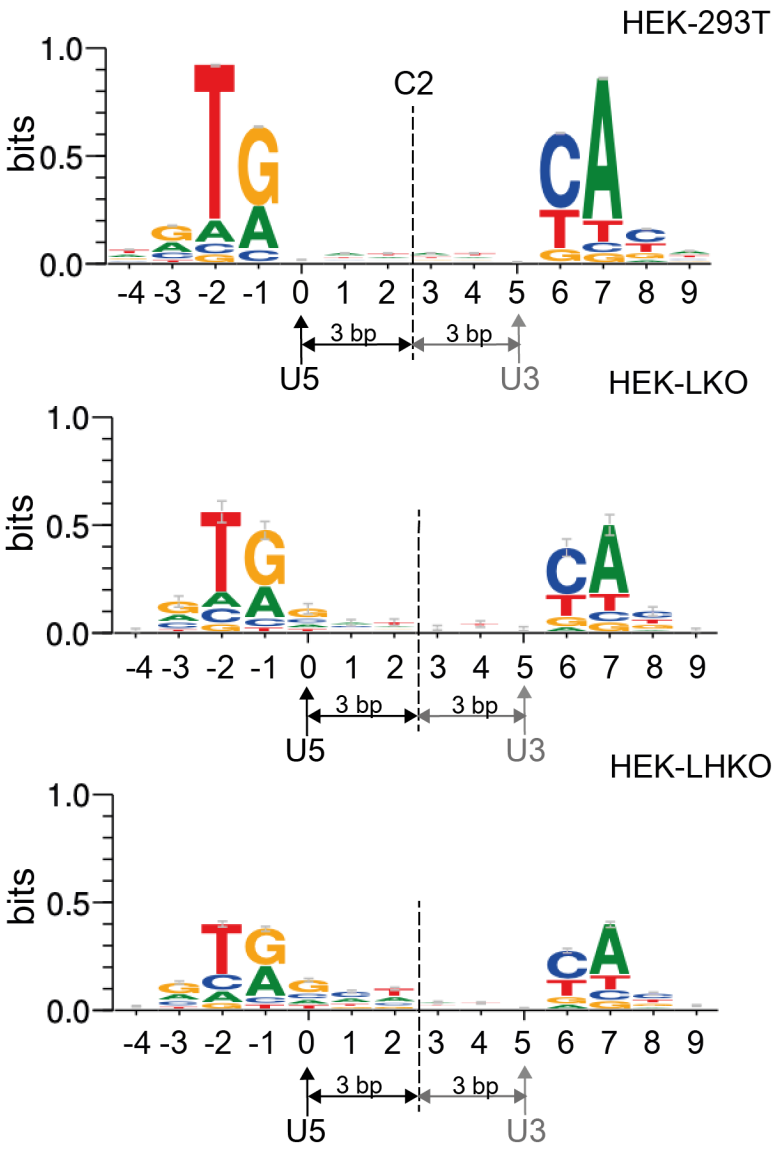

Figure 5-12 Nucleotide sequence preferences of MVV integration in HEK-293T cells.

Integration site sequence logos generated from HEK-293T, LKO, LHKO cells infected with WT MVV vectors. U5 vDNA end is joined to the target DNA at position 0. The U3 vDNA end is indicated in grey, with the two-fold symmetry (C2) indicated as a dashed line to form the symmetry axis.

5.6 Conclusions

LEDGF/p75 was essential for MVV IN strand transfer activity and intasome assembly *in vitro* (Figure 5-1). Both LEDGF/p75 and HRP2 were shown to potently stimulate HIV-1 IN activities *in vitro* (Cherepanov et al., 2003, Cherepanov et al., 2004, Hare et al., 2009b, Cook et al., 2020). However, under cell culture conditions, the effect of LEDGF/p75 and/or HRP2 depletion had variable effects on HIV-1 infectivity (Llano et al., 2004, Ciuffi et al., 2005, Llano et al., 2006, Shun et al., 2007). In all these studies, a consensus observation was made with respect to the shift in HIV-1 integration profiles in the absence of LEDGF/p75, which was exaggerated further in the absence of HRP2.

The dependence of MVV infectivity and integration targeting on LEDGF/p75 and/or HRP2 was investigated using LKO and LHKO cells. Initially, a strong dependence of MVV infectivity on the host factors was observed, wherein the infectivity of the WT MVV vector in LKO and LHKO cells was significantly lower than in the parental HEK-293T cells (Figures 5-3 and 5-4A), and this dependence was greatly reduced in the presence of the highest tested concentration (10,000 nM) of the INSTI, DTG (Figure 5-4A). However, a rescue phenotype of infectivity was not observed when ovine LEDGF/p75 was introduced into LHKO cells. This was attributed to the clonal variations that may have occurred in the LKO and LHKO clones, as well as the introduction of the ovine form of the protein (Figure 5-8). Although the human and ovine LEDGF/p75 sequences were 97% identical (Appendix figure 8.4), the possibility that the sheep protein failed to function in human cells cannot be ruled out. An unexpected effect of LEDGF/p75 and HRP2 was detected in cells infected with the IN enzyme active site mutant form of MVV (E154Q). The mutant was non-infectious parental HEK-293T cells, but a significant increase in luciferase expression was observed in LKO and LHKO cells, which was independent of the presence of INSTI drug (Figures 5-3 and 5-4B). This phenotype was partially reversed by the re-expression of LEDGF/p75 (Figure 5-8).

Sequencing of the integration sites from cells infected with MVV WT vector revealed both expected findings of shift in integration pattern (Table 5-1) as well as the surprising dependence of local nucleotide preferences on the host factors (Figure 5-12). This was unique to MVV integration in human cells and not observed in case of HIV-1 infection (Shun et al., 2007, Schrijvers et al., 2012a). MVV integration was generally targeted away from gene dense regions, without an obvious preference for highly expressive genes or an aversion for LAD regions. Absence of LEDGF/p75 during MVV infection also displayed an increased shift towards GC-rich regions in the genome, in comparison to that observed in HIV-1 infection.

These observations suggested that LEDGF/p75 might be playing an important role in not just the intasome assembly, but also ensuring its long-term stability in the cells infected with MVV. In its absence and without HRP2 to compensate for it, vDNA ends were not protected and are available for the host DNA repair machinery, causing an IN independent insertion of vDNA into the host genome. LEDGF/p75 and HRP2 also seem to be the key players in dictating the nucleotide preferences for MVV integration sites, which in case of HIV-1 were shown to be solely dependent on its intasome structure (Serrao et al., 2014). However, these observations cannot be assumed as conclusive. MVV is a sheep virus, with all the infections carried out in HEK-293T cells, and the observed phenotype was not reversed in the presence of *Oa*LEDGF/p75. These cells may lack ovine specific host factors that could stabilise the intasome in the absence of LEDGF/p75 and HRP2, corroborating with the observations made with HIV-1. The following chapter (Chapter 6) tests MVV infectivity in ovine cells knocked out for LEDGF/p75 and HRP2.

Chapter 6. Effect of LEDGF/p75 on MVV infectivity and integration site distribution in ovine cells

6.1 Aims

LEDGF/p75 was essential for MVV IN strand transfer *in vitro* and its depletion in human cells not only led to severe integration defects, but also caused significant redistributions of MVV integration sites (Chapter 5). Interestingly, an unexpected increase of MVV IN E154Q mutant's infectivity in the absence of LEDGF/p75 and more so in the absence of both LEDGF/p75 and HRP2 was detected, suggesting that these host factors played a critical role in stabilising the intasome. However, all these observations were in human cells and required verification in sheep cells, which would express ovine specific host factors that may be absent in human cells.

The aim for this part of the project was to test the importance of LEDGF/p75 for MVV intasome assembly in ovine cells and the necessity of the intasome formation for IN dependent integration of vDNA into host genome. LEDGF/p75 and HRP2 genes are encoded by *PSIP1* and *HDGF2* genes on sheep chromosomes 2 and 11, respectively. Both genes were disrupted in a sheep-derived cell line using CRISPR/Cas9 techniques, which allowed us to study MVV infection in the absence of LEDGF/p75 and HRP2.

6.2 Preliminary evaluation of ovine cells

Three clones of sheep choroid plexus cells, CPT-3, CPT-7, and CPT-8, immortalised with telomerase reverse transcriptase (TERT) (Arnaud et al., 2010) were obtained from Dr. David Griffiths, Moredun Research Institute, Midlothian, UK. Choroid plexus is part of the brain that produces cerebrospinal fluid and these cells are modified ependymal cells (Borlongan and Emerich, 2010). Firstly, the endogenous expression of LEDGF/p75 and HRP2 in these cell lines was tested by Western blotting (Figure 6-1).

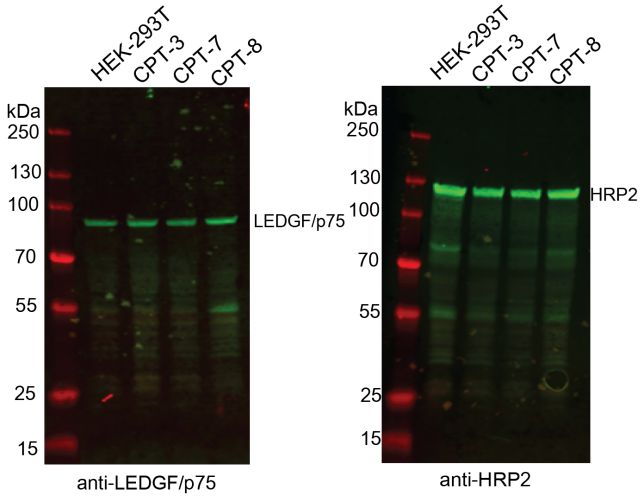

Figure 6-1 Detection of endogenous LEDGF/p75 and HRP2 in CPT cell lines.

Ovine CPT-3, -7 and -8 cells obtained from the Griffiths lab were tested for the presence of LEDGF/p75 and HRP2 by Western blotting. 10 µg (total protein) HEK-293T, CPT-3, CPT-7 and CPT-8 cell lysate was loaded per well and tested with anti-LEDGF/p75 or anti-HRP2 antibodies. Detection was carried out using near-infrared fluorescence.

Next, the cells were tested the ability to support infection with MVV WT and E154Q single cycle vectors. HEK-293T, CPT-3, -7 and -8 cells were seeded in 48-well plate and infected with 0.5 mU RT-normalised MVV WT or E154Q viruses. Infected cells were expanded 2 d post-infection into 6-well plates and harvested 7 d post-infection to measure luciferase expression (detailed method of infection in section 2.4.2).

HEK-293T cells were highly infectable with WT virus and E154Q displayed over a 100-fold defect for luciferase expression (Figure 6-2A). CPT clones were generally less infectable than HEK-293T cells, however, the difference in infectivity between WT and E154Q was unambiguously higher, with E154Q being 1,000-fold defective in comparison to WT infectivity (Figure 6-2B). WT infectivity was the highest in CPT-3 cells in comparison to CPT-7 and CPT-8 cells (Figure 6-2B). Subsequent experiments were carried out in CPT-3 cells.

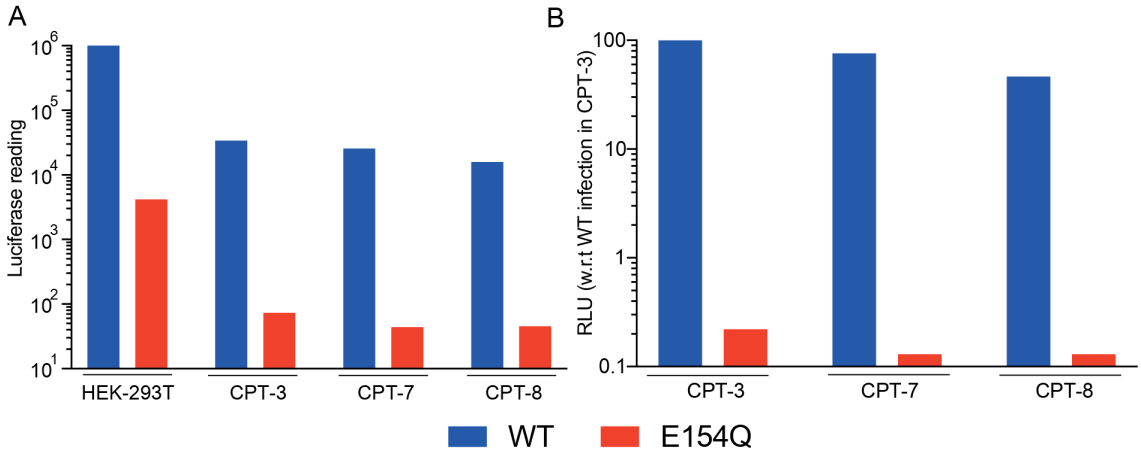

Figure 6-2 MVV vector infectivity measurement in CPT clones.

Infectivity of 0.5 mU RT-normalised MVV WT and E154Q viral vectors in HEK-293T, CPT-3, CPT-7 and CPT-8 cells was performed once. Luciferase was measured 7 d post-infection. (A) Luciferase measurements normalised to the total protein concentration, (B) Luciferase readings normalised to WT infectivity in CPT-3 cells.

6.3 Knockout of LEDGF/p75 in CPT-3 cells

6.3.1 CRISPR/Cas9 strategy for the knockout of sheep *PSIP1* gene

The IBD region of LEDGF/p75 was shown to be essential and sufficient for the interaction with lentiviral IN *in vitro* (Cherepanov et al., 2004, Cherepanov, 2007). In order to knockout sheep *PSIP1*, the region encoding the IBD was targeted. The corresponding region that encodes for the 83 amino acid residues, was analysed using Cas-OFFinder web server (Bae et al., 2014) for sequence repetitions in the ovine genome (*Ovis aries*, Oar_v3.1). Twenty-three bases in the IBD sequence 'TATTGAGGCTTTGGATGAACNGG' were found to be repetitive on sheep chromosomes 2 and 3. Therefore, this sequence was avoided during guide RNA (gRNA) design, to minimise the chances of unwanted genomic modifications. gRNAs (20 bases in length) were designed using the IDT custom gRNA design tool. gRNA sequences 1, 2, 6, 8 and 9 (highlighted in green) were outside the repetitive

sequence, from which 1, 2 and 6 were selected as gRNA 1, 2 and 3, gRNA sequences listed in Table 6-1.

The plus-strand oligos (F′) were designed with an upstream sequence of CACCG (5′-CACCGNNNNNNNNNNNNNNNNNNN-3′) and the minus strand oligos (R′) were designed with an upstream sequence of AAAC and a downstream C (5′-AAACNNNNNNNNNNNNNNNNNNNC-3′). The F′and R′oligos were annealed and ligated into *Bbs*I digested pSpCas9(BB)-2A-GFP, between the U6 promoter and sgRNA scaffold (Ran et al., 2013). The ligated plasmids, pSpCas9(BB)-2A-GFP-gRNA1/2/3 were confirmed by sequencing analysis.

Table 6-1 gRNA sequences targeting ovine *PSIP1* gene

	Position	Strand	Sequence	PAM	
1	224	+	CAAGTTTAAGAACATGTTTT	TGG	gRNA1
2	1	+	CAATGGATTCTCGACTTCAA	AGG	gRNA2
3	92	+	GGATGAACTGGCTTCACTTC	AGG	
4	80	+	TATTGAGGCTTTGGATGAAC	TGG	
5	71	+	CAACAGATGTATTGAGGCTT	TGG	
6	179	+	CAAAGTTAGTCAGGTTATCA	TGG	gRNA3
7	65	+	TGATGTCAACAGATGTATTG	AGG	
8	170	+	ACGGCGATTCAAAGTTAGTC	AGG	
9	151	+	TTACAACACTGAAAAAAATA	CGG	

6.3.2 Generation of clonal *PSIP1*-knockout CPT-3 cells

The transfection conditions for CPT-3 cells were optimised, by using two ratios of Xtreme gene 9 transfection reagent (µl) and DNA (µg) (3:1 or 6:1) (Sigma Aldrich). Addition of sodium butyrate (NaBu), an inhibitor of cellular histone deacetylases, was shown to increase the levels of transient protein expression (Parham et al., 1998). Transfections were carried out in the presence or absence of NaBu using the pSpCas9(BB)-2A-GFP-gRNA1/2/3 plasmids, that in addition to the Cas9 and gRNA expression cassette encode GFP, which allowed to monitor transfection efficiency. Cells were tested for GFP expression, 2 and 3 d post-transfection. Cells transfected with the higher amount of transfection reagent to DNA ratio (6:1) required an additional day to recover and attain near confluency (Figure 6-3). However, under all

conditions tested, no significant differences in GFP expression were observed (indicated for gRNA1, Figure 6-3).

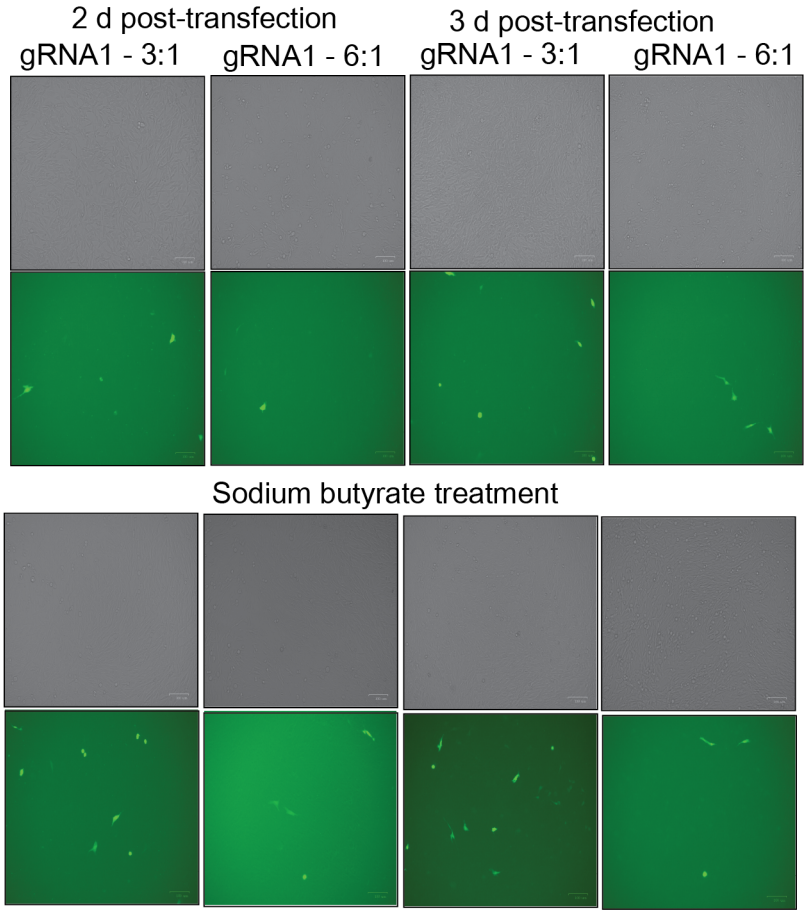

Figure 6-3 Transfection optimisation.

CPT-3 cells were transfected with pSpCas9(BB)-2A-GFP-gRNA1/2/3 plasmids with the indicated ratio of transfection reagent and DNA, in the presence or absence of sodium butyrate. GFP expression was analysed 2 and 3 d post-transfection using Zoe fluorescent cell imager (Bio-Rad). Only images of cells transfected with pSpCas9(BB)-2A-GFP-gRNA1 shown.

CPT-3 cells were transfected with the gRNA encoding plasmids, in the absence of NaBu and harvested for single cell sorting by flow cytometry. Cells were resuspended in FACS buffer at $4x10^6$ cells/ml and single GFP-positive cells were sorted using a 100 µm nozzle on a MoFlo XDP High-Speed Cell Sorter at the Flow Cytometry STP (Francis Crick Institute). Single GFP-positive cells were deposited (Figure 6-4) into individual wells of a 96-well plate containing pre-conditioned media. Non-transfected CPT-3 cells were used as GFP-negative control cells. The methods of transfection and cell sorting are detailed in sections 2.4.3 and 2.6.2.

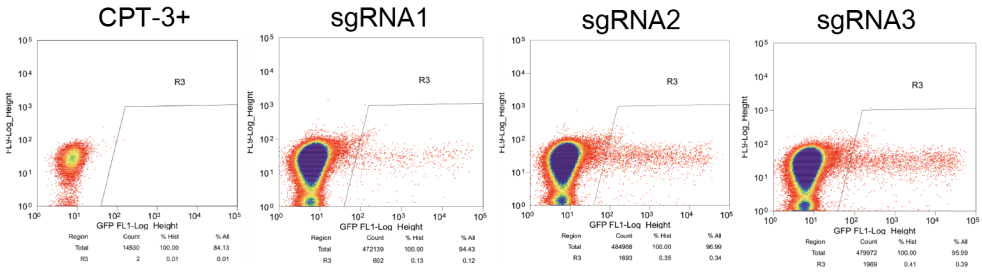

Figure 6-4 Flow sorting of transfected cells.

Single GFP-positive cells were sorted by flow cytometry. Non-transfected CPT-3 cells were used as GFP-negative control.

The cells were initially expanded into 12-well plates, 21 days post-sorting and subsequently expanded into 6-well plates. Once confluent, they were harvested and whole-cell lysates tested for analysed by Western blotting with anti-LEDGF/p75 antibody (Appendix figure 8-5A). Samples 1 to 29 were from gRNA1, 30 to 44 from gRNA2 and 45 to 59 from gRNA3. Clonal cell lines 46, 54, 55, 56 and 57 were expanded further and re-tested for LEDGF/p75 expression (Appendix figure 8-5B). Consequently cell clones 46, 55 and 56 were selected as CPT-3 LKO cells (Figure 6-5).

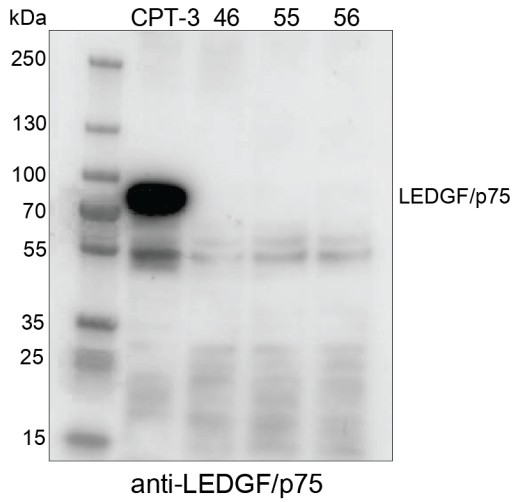

Figure 6-5 Testing LEDGF/p75 expression of CPT-3 LKO cells.

Selected clonal cell lines, generated from sgRNA3 were tested for expression of LEDGF/p75 by Western blotting. 20 µg (total protein) of CPT-3, LKO 46, LKO 55 and LKO 56 cell lysate was loaded per well and tested with anti-LEDGF/p75 antibodies. Detection was carried out by measuring chemiluminescence, using a secondary antibody tagged with HRP.

6.3.3 Sequence analysis of CPT-3 LKO cells

LEDGF/p75 expression was not detected by Western blotting in CPT-3 LKO clones, 46, 55 and 56, which were all generated from cells transfected with gRNA3. Their genomic regions surrounding the gRNA3 target sequence were analysed to map the specific locations of frameshifts or deletions. Sequencing primers were designed based on the ovine *PSIP1* gene sequence (RefSeq: NM_001031610.2) and PCR reactions were carried with genomic DNA purified from the cells. PCR reaction products were ligated into pCR-TOPO-TA vector (Life Technologies), and plasmid DNA purified from twelve single bacterial colonies was submitted for sequencing, as detailed in section 2.2.9.

Sequence analysis revealed that LKO 46 cells had a 64 bp deletion, effectively removing part of the LEDGF/p75 open reading frame within of IBD α-helix 5. LKO 55 and 56 cells had multiple single- and double-bp deletions, also disrupting the coding

sequence just within α-helix 5. The residues of the LEDGF/p75 IBD, ^{364}KIDNLDV^{370} (indicated in red, Figure 6-6), responsible for interacting with the IN CCD are present in α-helices 1 and 2, discussed in section 1.5.6, Figure 1.14. Interfering with these residues was shown to completely abrogate IN-LEDGF interaction (Cherepanov et al., 2005b).

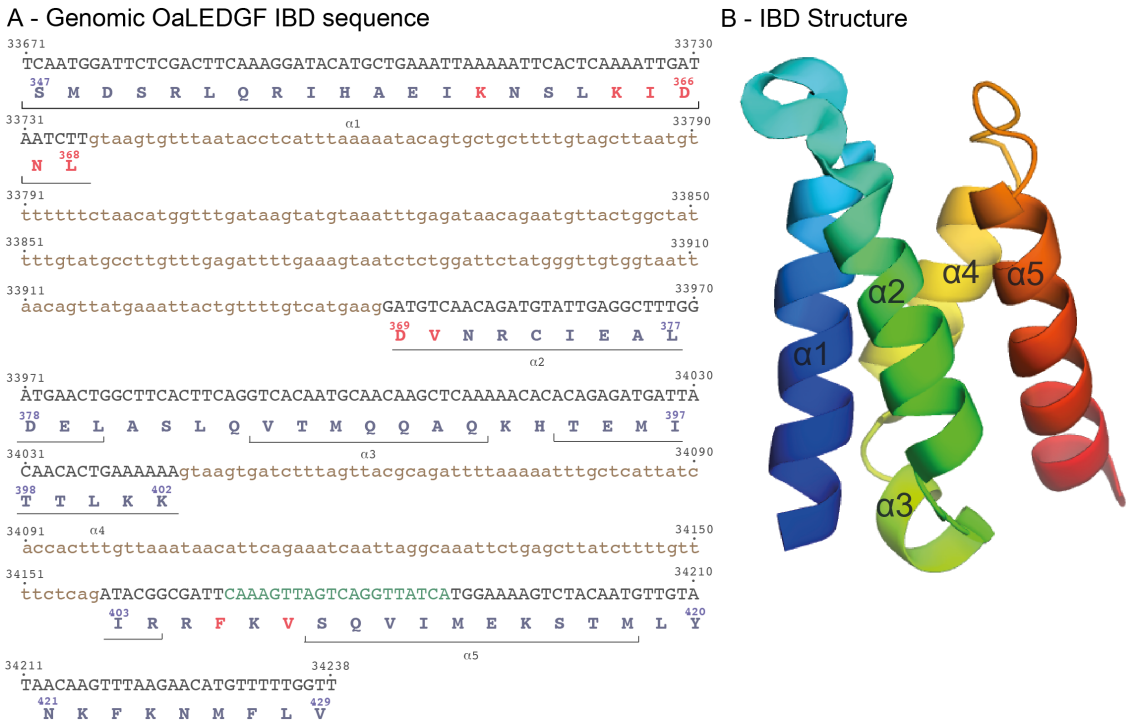

Figure 6-6 Sheep genomic sequence surrounding the region encoding OaLEDGF/p75 IBD.

(A) The IBD-encoding region on ovine chromosome 2 (33671 to 34238), with the intron sequence in lower case (brown) and encoded amino acid encoding exons in upper case (black). The IBD amino acid sequence (347 to 429, numbered from 1-530 in LEDGF/p75) and α-helices indicated in blue. gRNA3 sequence indicated in green. (B) Structure of IBD (Protein Data Bank ID 1Z9E) with the five α-helices indicated.

6.4 Nucleofection of CRISPR/Cas9 complex to generate CPT-3 LKO and LHKO cells

The only gRNA construct that led to a deletion within the region encoding LEDGF/p75 IBD region was targeted towards its C-terminus, within α-helix 5 (Figure 6-6). Unfortunately, the genomic modification may have preserved the most important region of the LEDGF/p75 open reading frame. The reduced LEDGF/p75 expression in these clones is probably explained by destabilisation of the protein product due to disruption of IBD folding. However, there was a possibility of minor traces of LEDGF/p75 with IBD largely intact present in these cells. Indeed, HIV-1 infection was not significantly impaired in cells with only a knockdown of LEDGF/p75 (Llano et al., 2004). Another potential pitfall is the clonal nature of the resulting cell lines and with the transfection efficiency and GFP expression being relatively low in CPT-3 cells, nucleofection was used as an alternative method. Moreover, to improve the efficiency of genome modification, pre-formed CRISPR/Cas9 ribonucleoprotein (RNP) complexes were used (Kim et al., 2014).

6.4.1 Optimisation of Nucleofection

The parameters for nucleofection were optimised using the Cell Line Optimization 4D-Nucleofector X Kit (Lonza). The electroporation buffer conditions and the nucleofection pulse rate and voltage, from 15 Nucleofector programs were determined (Figure 6-7) by transfection of pmaxGFP reporter plasmid, supplied with the kit, process elaborated in section 2.6.7. Cells were harvested 2 d post-nucleofection and tested for GFP expression using flow cytometry. The program setup of E2 (EN-138), with the electroporation buffer SE were determined as the ideal conditions for nucleofecting CPT-3 cells.

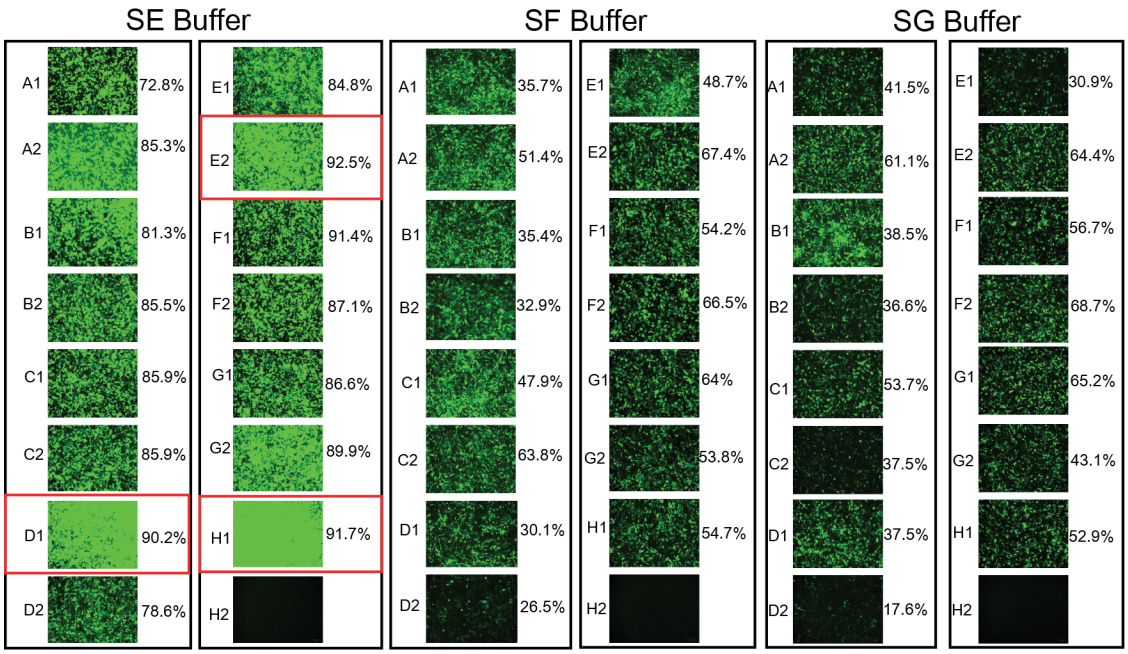

Figure 6-7 Optimisation of nucleofection parameters for transfection of CPT-3 cells.

The parameters for carrying out nucleofection were optimised using 15 nucleofection settings, with three electroporation buffers (SE, SF and SG). GFP expression was analysed 2 d post-nucleofection using Zoe fluorescent cell imager (Bio-Rad). The percentage of GFP-positive cells measured by flow cytometry are indicated for each condition.

6.4.2 Design of crispr RNA (crRNA) for targeting sheep *PSIP1* and *HRP2* genes

The crRNA sequences were designed using ovine *PSIP1* (RefSeq: NM_001031610.2) and *HDGF2* (RefSeq: NM_001114761.1) gene sequences (Kent et al., 2002), detailed in Table 6-2. The sequence of exons and introns indicated in upper and lower cases, respectively.

Table 6-2 Details of crRNA sequences.

		Sequence of crRNA	Position on the chromosome	
1	crRNA_1	ttaatttatagAAACATCAA	*Oa* LEDGF IBD	33655 to 33674
2	crRNA_2	CAATGGATTCTCGACTTCAA	*Oa* LEDGF IBD	33672 to 33691
3	crRNA_3	CAACAGATGTATTGAGGCTT	*Oa* LEDGF IBD	33948 to 33967
4	crRNA_4	TATTGAGGCTTTGGATGAAC	*Oa* LEDGF IBD	33957 to 33976
5	crRNA_5	cctacctgcagAGCCCTCCG	*Oa* HRP2 IBD	13516 to 13535
6	crRNA_6	acctgcagAGCCCTCCGTGG	*Oa* HRP2 IBD	13519 to 13538
7	crRNA_7	tcacCGGATTGTCAACCTTC	*Oa* HRP2 IBD	13582 to 13601
8	crRNA_8	TGAGATCAAGTTCGCCCTGA	*Oa* HRP2 IBD	13564 to 13583
9	crRNA_9	GTATTACTGATATTGGTGGG	Negative control	

6.4.3 Nucleofection of CPT-3 cells

Equimolar ratios of crRNA and tracrRNA (universal 67-mer, Alt-R CRISPR-Cas9 tracrRNA, IDT DNA) were hybridised to form gRNA and ribonucleoproteins (RNPs) were formed by incubating gRNA and recombinant Alt-R S.p. Cas9 Nuclease (IDT DNA). CPT-3 cells were resuspended in SE electroporation buffer and added to the nucleofection chamber, along with the RNPs. The cells were electroporated with the EN138 program and the treated cells were plated in 6 well plates. The detailed method of nucleofection is elaborated in section 2.6.8.

Cells were harvested 3 d post-nucleofection and analysed for LEDGF/p75 and HRP2 expression by Western blotting (Figure 6-8). As expected, transfection with the control RNP did not affect LEDGF/p75 or HRP2 expression. By contrast, the RNPs targeting LEDGF/p75 IBD (formed using gRNAs 1 - 4) severely depleted LEDGF/p75 expression. Knockout of HRP2 expression with RNPs formed with gRNAs 5 - 8 was

less efficient, although a noticeable reduction in HRP2 levels was achieved, while gRNA9 was ineffective (Figure 6-8).

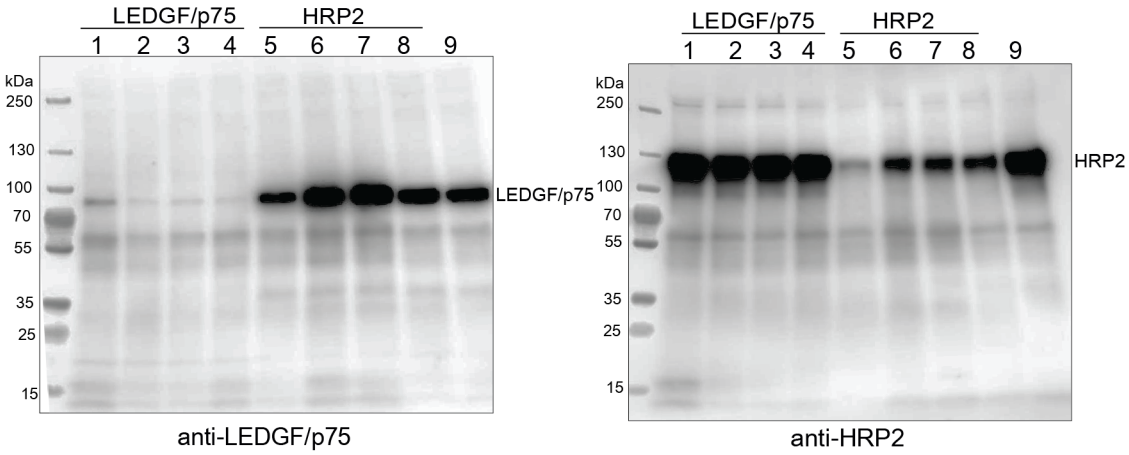

Figure 6-8 Initial nucleofection of CPT-3 cells with gRNAs targeting sheep *PSIP1* and *HRP2* genes.

Nine nucleofections were performed with the RNPs in CPT-3 cells and protein expression analysed 2 d post-nucleofection by Western blotting. CPT-3 cells nucleofected with gRNAs targeting LEDGF/p75 (lanes 1 to 4), gRNAs targeting HRP2 (lanes 5 to 8) and the control gRNA (lane 9).

Subsequent nucleofections of sheep CPT-3 cells were carried out using combinations of RNPs. Four *PSIP1* knockout (LKO) polyclonal cell lines were generated, LKO1 (transfected with a combination of RNPs 2, 3, and 4), LKO2 (RNPs 2, 4, and 3), LKO3 (RNPs 2 and 3) and LKO4 (RNPs 2 and 4). In addition, two *PSIP1/HRP2* double knockout (LHKO) cell lines were attempted, LHKO1 (RNPs 2, 3, 5, 6, 7, and 8) and LHKO2 (RNPs 2, 4, 5, 6, 7, and 8). Western blotting of cell extracts prepared 7 days post-nucleofection revealed a highly efficient depletion of LEDGF/p75 expression, while the expression of HRP2 was severely reduced in both LHKO cell lines (Figure 6-9).

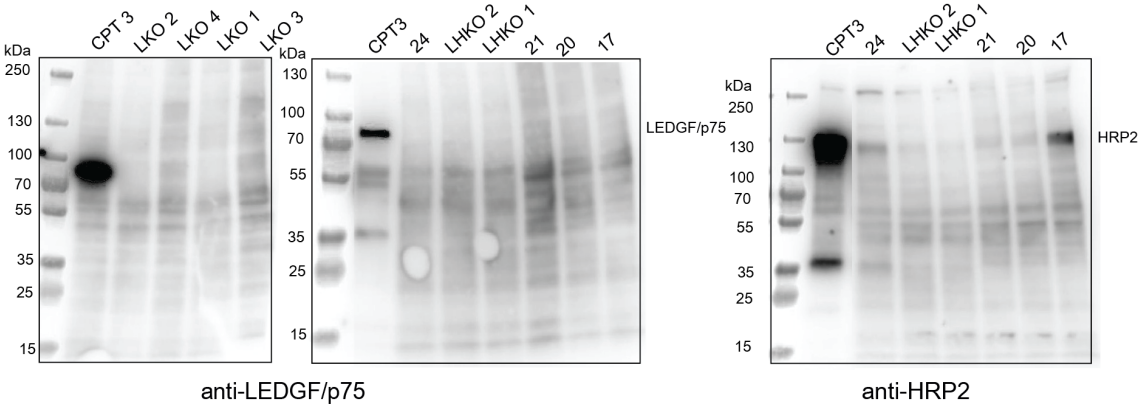

Figure 6-9 Generation of CPT-3 LKO and LHKO cells.

Cell lines grown out after nucleofection with gene targeting RNPs were tested for LEDGF/p75 and HRP2 expression by Western blotting. Four LKO cell lines (LKO 1 to LKO 4) and two LHKO cell lines (LHKO 1 and LHKO 2) were selected for further experiments.

6.5 Infections in CPT-3 LKO and LHKO cells

CPT-3, clonal LKO 46, LKO 55, LKO 56 cells and nucleofected LKO 1, LKO 2, LKO 3, LKO 4, LHKO 1 and LHKO 2 cells were seeded in 48-well plates and infected with 0.5 mU RT-normalised MVV WT or E154Q vectors. 2 d post-infection, cells were expanded into 6-well plates and harvested 7 d post-infection to measure luciferase activity. Infected cells were treated as described in section 2.6.3 and luminescence was measured.

WT infectivity in CPT-3 cells was normalised as 100% and as expected, E154Q vector was nearly 1000-fold defective for infectivity in CPT-3 cells in comparison to WT (Figure 6-10). Interestingly, both WT and E154Q infectivity in the clonal LKO cells, 46 and 55 was higher than the parental CPT-3 cells. A minor decrease in WT infectivity was observed in LKO 56 cells, with a minor increase on E154Q infectivity. A significant reduction in WT infectivity was not observed in all the nucleofected polyclonal LKO and LHKO cells and the infectivity of E154Q was comparable to the

parental CPT-3 cells, in exception to LKO 2 cells, wherein a significant 10-fold increase was observed (Figure 6-10).

An increase in infectivity LKO 46 and 55 cells and the lack of a significant difference in LKO 56 cells was attributed to the presence of HRP2, which could highly compensate for the absence of LEDGF/p75 in MVV infection or a consequence of the clonal variation, in comparison to the parental CPT-3 cells (Lee et al., 2019). The WT infectivity in LEDGF/p75 and HRP2 depleted CPT-3 cells were in stark contrast to the observations made in HEK-293T LKO and LHKO cells (Figure 5-3).

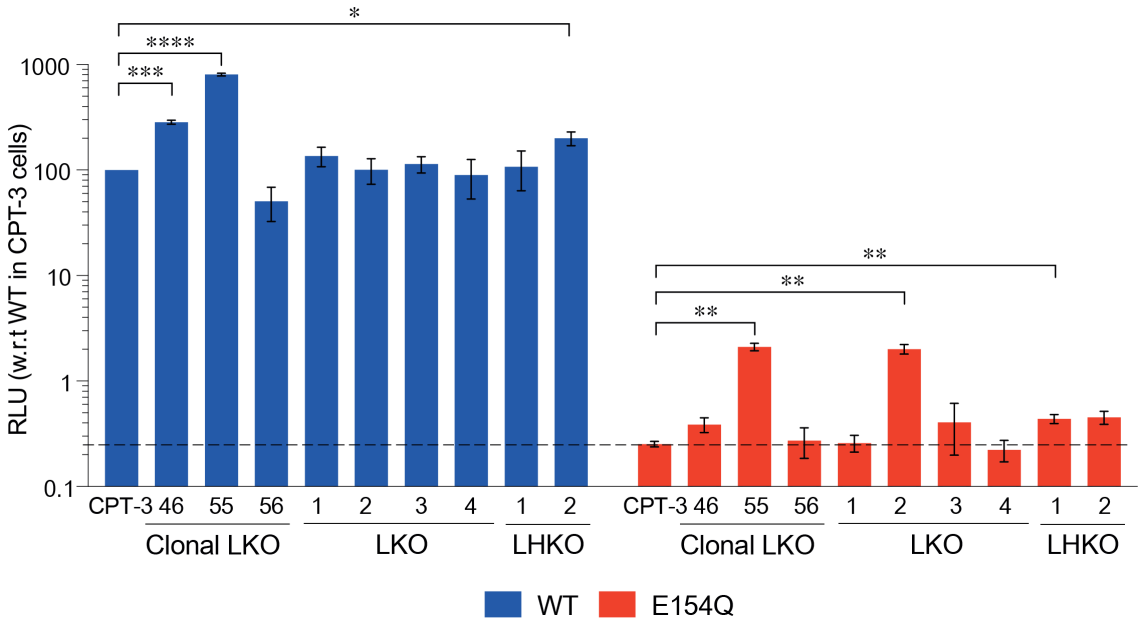

Figure 6-10 MVV vector infectivity in CPT-3 LKO and LHKO cells.

CPT-3, LKO and LHKO cells infected with 0.5 mU RT-normalised MVV WT or E154Q viral vectors and luciferase measured 7 d post-infection. RLUs were normalised to WT infectivity in CPT-3 cells, which was set to 100%. Means with standard error of the mean plotted of three biological repeats, with two technical repeats each; Two-tailed paired t-tests (individual comparisons), p=0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****), with only significant differences indicated.

6.6 Testing the presence of residual LEDGF/p75 in CPT-3 LKO and LHKO cells

The experiments revealed a modest effect of LEDGF/p75 and LEDGF/p75/HRP2 depletion on the transduction of sheep cells with WT MVV vector and variable effect on the infectivity of E154Q MVV vector. These observations were in contrast to the results obtained with the clonal human LKO and LHKO cells. One possible explanation is that the sheep LKO and LHKO cells retained small but significant residual expression of LEDGF/p75. It was previously shown that residual protein maybe present associated with chromatin in cells knocked down with siRNAs targeting LEDGF/p75, which can cause lack of expected phenotypes (Llano et al., 2006). Therefore, a previously described protocol for subcellular fractionation was used to enhance the detection of LEDGF/p75. $10x10^6$ cells each of the cell lines were lysed and the total (T) fraction, and Triton-X100-soluble protein fraction (S1) were separated. The remainder was DNase- and salt- treated, to isolate the DNase-released chromatin-associated fraction (S2). Presence of LEDGF/p75 in these fractions was tested by Western blotting.

As expected, no LEDGF/p75 expression was detected in the total fraction (T) in all the KO cells generated. Most of LEDGF/p75 protein in CPT-3 cells was associated with chromatin, which was only released by DNase treatment (Figure 6-11). The clonal cell lines, LKO 46, LKO 55 and LKO 56 and nucleofected LKO 1 and LKO 2 cells did not have detectable levels of LEDGF/p75 in both S1 and S2 fractions. Scant amounts of LEDGF/p75 was present associated with chromatin in LKO 3, LKO 4, LHKO 1 and LHKO 2 cells (Figure 6-11).

Although variable effects of LEDGF/p75 and HRP2 depletion were observed in CPT-3 cells infected with MVV WT and E154Q, these observations do not corroborate with the observations made in HEK-293T LKO and LHKO cells. While the possibility of presence of other ovine specific host factors cannot be ruled out, the observations presented here suggest that LEDGF/p75 and HRP2 maybe dispensable for MVV infectivity.

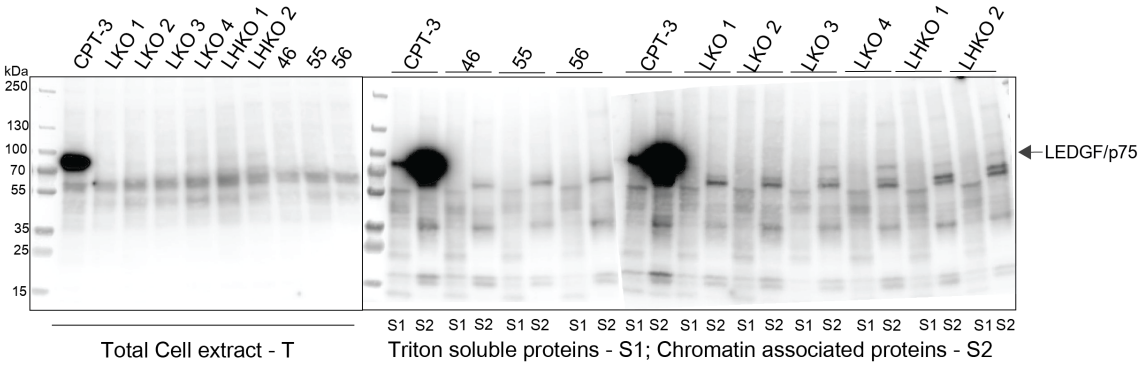

Figure 6-11 Enhanced detection of residual LEDGF/p75 expression using cell fractionation.

CPT-3, clonal LKO, and polyclonal LKO and LHKO cell lines were tested for the presence of LEDGF/p75 associated with chromatin. Ten-million cells each of the cell lines were lysed, triton soluble (S1) and chromatin associated proteins (S2) were separated from the total cell extract (T). 20 µg (total protein) separated fractions were loaded per well and tested with anti-LEDGF/p75 antibodies. Detection was carried out by measuring chemiluminescence, using a secondary antibody tagged with HRP.

6.7 Integration site sequencing

The integration pattern of MVV infection in ovine CPT-3 cells knocked out for LEDGF/p75 and HRP2 was determined by integration site sequencing, elaborated in section 5.5. CPT-3, LKO 1 and LHKO 1 and LHKO 2 cells were infected with fresh MVV WT viral preparations and infected cells were harvested 5 d post-infection. DNA was purified from the infected cells and samples were processed further by Dr. Parmit Singh (Engelman laboratory), as described in section 5.5. The samples were analysed by Illumina sequencing (Genewiz Laboratories) and the sequencing data was processed by Dr. Parmit Singh and Prof. Peter Cherepanov, (described in Serrao et al., 2016) and integration site maps were generated.

CPT-3 cells infected with WT virus displayed 72.1% integration events targeted into the genes, while the integration events in genes in LKO 1 (51.8%), LHKO 1 (47.6%) and LHKO 2 (51.6%) cells were closer to the MRC values (43.3%) (Table 6-3). A 2 to 3-fold increase in integration events at the vicinity of TSSs (2.5 kb of TSS) and CpG islands was observed in LKO and LHKO cells (Table 6-3), which is in accordance with previously reported observations with HIV-1 infection (Wang et al., 2012, Sowd et al., 2016, Li et al., 2020).

Table 6-3 MVV integration site distributions in CPT-3 cells, LKO and LHKO cells

Virus or control[a]	Cells	Unique sites	RefSeq genes (%)[b]	±2.5 kb TSS (%)[c]	±2.5 kb CpG (%)[c]	Gene density[d]
MRC	*in silico*	284,737	43.3	4.6	6.1	8.1
MVV	CPT3	296,801	72.1	5.2	4.4	10.7
	CPT3-LKO1	145,863	51.8	12.4	16.5	12.7
	CPT3-LHKO1	3,880	47.6	10.9	14.9	12.2
	CPT3-LHKO2	82,570	51.6	12.4	16	12.2

[a] Matched random control (MRC) or a single-cycle MVV GFP vector.

[b] Integration events inside a genomic feature (RefSeq gene).

[c] Integration events within 2.5 kb of a feature (TSS or CpG island).

[d] Average number of RefSeq genes within a 1-Mbp window (±500 kbp) of integration site.

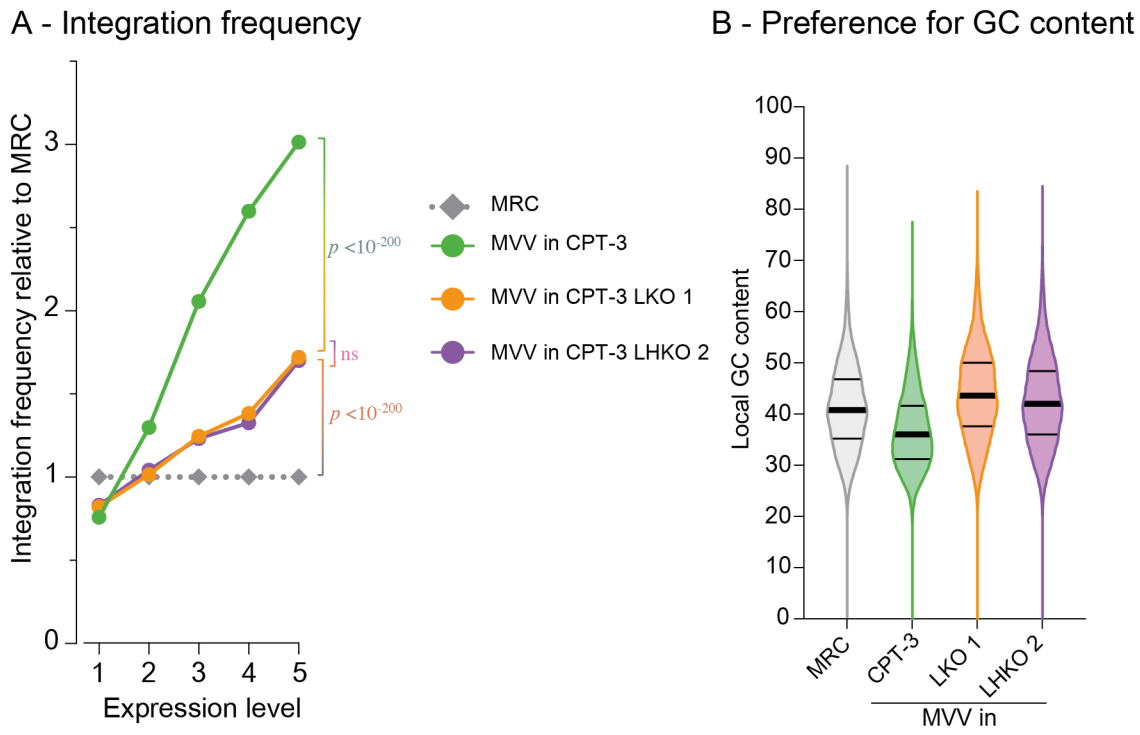

Figure 6-12 MVV integration site preferences in CPT-3, LKO and LHKO cells.

(A) Sheep genes were divided between 5 bins based on their expression level in ovine brain tissue (bins 1 and 5 contained genes with the lowest and highest expression, respectively). MVV integration sites were scored in each bin. Note the strong preference of MVV to integrate into more active TUs. By contrast, in silico-generated matched random control (MRC) integration sites shows equiprobable distribution between bins. (B) Genomic GC content within close proximity of integration sites (within 500 bp windows) is shown in form of a violin plot for each sample. Thick black lines indicate median values, and thin black lines demarcate quartiles.

MVV integration frequency was calculated in genes based on their expression levels in the brain tissues of sheep (Clark et al., 2017), as described in section 5.5.2. The frequency of MVV integration into highly expressed genes was considerably higher in sheep cells, than in HEK-293T cells and this preference was retained in both the KO cell lines (Figure 6-12A). Although no significant reduction in MVV infectivity was observed in CPT-3 LKO cells, with an apparent increase in infectivity observed in

LHKO cells (Figure 6-10), intriguingly a prominent shift in integration targeting towards GC-rich regions was observed in these cells (Figure 6-12B). Furthermore, the local nucleotide sequence preference in the CPT-3 KO cells was similar to the parental CPT-3 cells (Figure 6-13), without the significant decrease that is observed in human LKO and LHKO cells (Figure 5-12).

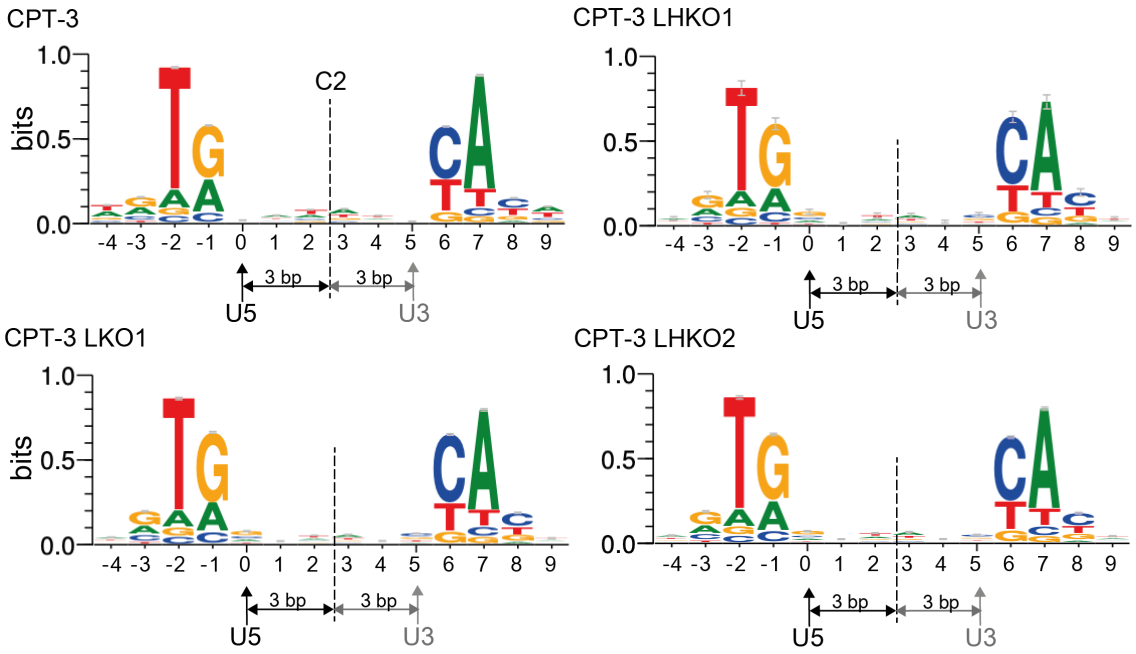

Figure 6-13 Nucleotide sequence preferences of MVV integration in CPT-3 cells.

Integration site sequence logos generated from CPT-3, LKO 1, LHKO 1 and 2 cells infected with WT MVV vectors. U5 vDNA end is joined to the target DNA at position 0. The U3 vDNA end is indicated in grey, with the two-fold symmetry (C2), is indicated as a dashed line to form the symmetry axis.

6.8 Conclusions

LEDGF/p75 played a key role in assembling the MVV intasome in infected cells and the intasome assembly was necessary to carry out IN dependent integration of vDNA ends into the host genome (Chapter 5). These observations were unique to MVV integration in HEK-293T LKO and LHKO cells. The importance of LEDGF/p75 for MVV integration in ovine cells was tested. Ovine CPT-3, -7, -8 cells expressed detectable levels of both LEDGF/p75 and HRP2 proteins (Figure 6-1). Preliminary infection of these cells revealed a 1000-fold defect in the luciferase expression of cells infected with E154Q, with respect to WT infection, with CPT-3 cells being the most infectable cells (Figure 6-2).

In order to knockout LEDGF/p75 from CPT-3 cells, CRISPR/Cas9 based methods were adopted. gRNAs targeting *Oa*LEDGF/p75 IBD were introduced into CPT-3 cells and single GFP-positive cells were sorted using flow cytometry (Figure 6-4). Three clonal cell lines were determined as LKO cells, LKO 46, 55 and 56 (Figure 6-5). However, sequencing the KO cells revealed a deletion downstream of the α-helix 5 of the IBD. Nucleofection of the CRISPR/Cas9 complexes was used as an alternative method to knockout LEDGF/p75 and HRP2 to generate LKO and LHKO cells. gRNAs targeting the N-terminal region of the *Oa*LEDGF/p75 were introduced into CPT-3 cells to generate four LKO cells, LKO 1, LKO 2, LKO 3 and LKO 4, followed by gRNAs targeting the N-terminal region of the *Oa*HRP2 IBD, to generate LHKO 1 and LHKO 2 cells (Figure 6-9).

WT infectivity was not severely impaired in all the sheep LKO and LHKO cells generated. In fact, an increase in infectivity was observed in clonal LKO 46, 55 cells and polyclonal LHKO 2 cells (Figure 6-10). Presence of residual levels of LEDGF/p75 were observed associated with the chromatin in LKO 3, LKO 4 and both the LHKO cells (Figure 6-11). Although the lack of expected phenotypes observed in the clonal cells can be attributed to their clonal nature, the complete lack of predicted phenotype in the polyclonal KO cells, especially in LKO 1 and LKO 2 cells, points either towards experimental inaccuracies or the presence of ovine specific host factors that highly compensate for the absence of LEDGF/p75 and HRP2.

LKO 1 and LHKO 1 and LHKO 2 cells were infected with WT MVV vectors and integrations sites were mapped. In spite of the presence of residual levels of LEDGF/p75 in the LHKO cells, a significant shift in integration pattern was observed in the KO cells (Table 6-3). MVV integration strongly favours integration into genes, similar to the other lentiviruses, but unlike HIV-1, -2, and similar to BIV and FIV, its preference for highly dense gene regions is not a strict requirement and it is not dictated by the presence of LEDGF/p75 (Li et al., 2020). It is interesting to note that the frequency of integration into highly expressed genes is higher (Figure 6-12A), as well as the local integration site preferences were restored in sheep cells (Figure 6-13), in contrast to the observations made in human cells (Figures 5-11A and 5-12). This preference is not completely lost in the absence of the host factors (Figure 6-12A), further pointing to the presence of ovine specific host factors that aid in targeting integration towards highly transcribing genes. However, consistent with the role played by LEDGF/p75 in interacting with the A/T rich regions of the genome, an increase in integration towards GC-rich regions was observed in both the KO cell lines tested (Figure 6-12B).

Variable levels of luciferase expression were observed in CPT-3 LEDGF/p75 and HRP2 depleted cells infected with E154Q, with all the KO cell lines generated, in exception to LKO 55, LKO 2 and LHKO 1, did not exhibit a significant increase in luciferase expression (Figure 6-10). The apparent increase in E154Q infectivity in the clonal LKO 55 and polyclonal LHKO 2 cells needs to be disregarded, due to a general increase in infectivity of these cells, as observed with WT vectors. A significant increase in infectivity was only observed in LKO 2 and LHKO 1 cells (Figure 6-10).

The observations reported here require further evaluation, especially with the presence of HRP2 in the clonal LKO cells and only a knockdown of HRP2 in nucleofected LHKO cells. Interesting insights into MVV integration were gained by sequencing the integration sites in these KO cells. Although the general lentiviral preferences for integration sites were observed, MVV specific features were also observed. These MVV specific features were more prominent in sheep cells and not observed in human cells. However, all the observations reported above undoubtedly reinforce the importance of LEDGF/p75 for MVV integration.

Chapter 7. Discussion

Retroviral integration is an essential step in establishing viral infection. The two catalytic steps involved in viral integration, 3′-processing and strand transfer, are performed by viral IN. Retroviral INs assemble into homomultimers in complex with vDNA ends and, in some cases, have been shown to recruit host factors. The PFV, HTLV-I and STLV intasomes were characterised to harbour tetrameric IN assemblies (Hare et al., 2010, Bhatt et al., 2020, Barski et al., 2020), whereas IN octamers were observed in the intasomes from alpha- and betaretroviruses (Yin et al., 2016, Ballandras-Colas et al., 2016, Pandey et al., 2021). The lentiviral intasomes were the largest complexes observed in the family. Thus, the MVV structure featured a hexadecameric IN assembly (Ballandras-Colas et al., 2017), while a presumably partial HIV-1 intasome structure harboured a dodecamer (Passos et al., 2017). Despite this unexpected diversity, all retroviral intasomes characterized to-date contained a CIC, which supports a pair of active sites for catalysing the IN-mediated reactions of 3′-processing and strand transfer.

The MVV intasome is assembled from four IN tetramers, organised with two catalytic tetramers that contribute one active site each, plus two flanking tetramers that contribute the synaptic CTDs to complete the CIC (Ballandras-Colas et al., 2017). The dodecameric HIV-1 intasome contains two catalytic IN tetramers and two flanking dimers, thus missing a dimer from each flanking IN tetramer (Passos et al., 2017). By contrast, in spumaviruses and deltaretroviruses that contain the minimal tetrameric IN complex, consisting of little more than the CIC, a tetramer is sufficient to perform both the catalytic steps involved in the integration reaction (Maertens et al., 2010, Hare et al., 2012, Barski et al., 2020, Bhatt et al., 2020). The questions that arise are: (*i*) why the other intasome assemblies need to be 2- to 4-fold larger than the minimalistic assemblies observed in PFV, HTLV-I, and STLV intasomes and (*ii*) is it possible to assemble a functional lentiviral intasome *in vitro* from a reduced number of IN subunits, and whether it would be able to carry out integration during viral infectivity. The latter was addressed in this study, in particular the importance of the lentiviral intasome assembly in the context of viral infection and the functional consequences of the extended architecture the MVV intasome were examined.

7.1 Key findings

7.1.1 Structure-based site-directed mutagenesis of MVV IN

The importance of the hexadecameric assembly within the MVV intasome was tested by introducing amino acid substitutions in IN, targeting its lentivirus-specific features: the CTD tetrad bringing the two halves of the intasome together and the α-helical CCD-CTD linker. Importantly, the CTD-CTD interactions and the α-helical configuration of the CCD-CTD linker are conserved in and unique to lentiviral INs (Ballandras-Colas et al., 2017) (Appendix figure 8-6). Single amino acid substitutions were introduced to interfere with the inter- (F223A, Y225A and I272E) and intra- (W245E/L, V252A/D, Y261A and V263E) tetramer interactions observed in the CTD bridge (Table 7-1). Likewise, mutations were introduced to destabilise the α-helical CCD-CTD linker (QQ207GP, QQQ211PGG and QQQ211PFVlinker). Through the CTDs, the linkers also contribute to the intra-tetramer interactions (Table 7-1). The intra-tetramer interactions were also observed in both HIV-1 STC and SIV intasome structures (Passos et al., 2017, Cook et al., 2020), spanning over the amino acid residues Trp-243 to Pro-261 (Figure 7-1). It is interesting to note that although the CTD is the most variable IN domain, the hydrophobic patches between MVV IN amino acid residues Val-243 to Trp-245, Ala-250 to Val-253 are conserved among the lentiviral integrases analysed (Appendix figure 8-6). Importantly, the amino acid residues targeted within MVV IN do not directly participate in the CIC formation, however, some of the CTD residues targeted in this study are involved in contacts with bases of non-transferred vDNA strand. All known interactions of for the residues substituted in this work are listed in Table 7-1.

Table 7-1 Interactions of targeted residues within the intasome.

MVV IN residue	Mutation	Structurally equivalent HIV-1 IN residues	Position in the intasome	Potential intasome interactions
Phe-223	F223A	Phe-223	CTD bridge	inter-teramer
			CTD Bridge	from CTD bridge hydrophobic interactions in the same tetramer
Tyr-225	Y225A	Tyr-226	CTD bridge	inter-teramer
			CTD Bridge	from CTD bridge hydrophobic interactions in the same tetramer
Trp-245	W245E/L	Trp-243	CTD Bridge	intra-tetramer
			Catalytic tetramers	NTD of same IN
			Synaptic CTDs	non-transferred strand of vDNA
Val-252	V252A/D	Val-250	CTD Bridge	intra-tetramer
			Catalytic tetramers	NTD from same IN
			Synaptic CTDs	non-transferred strand of vDNA
Tyr-261	Y261A/E	Ile-257	CTD Bridge	intra-tetramer
			Catalytic tetramers	NTD of same IN
			Synaptic CTDs	non-transferred strand of vDNA
Val-263	V263E	Val-260	CTD bridge	intra-tetramer
			Catalytic tetramers	NTD from same IN
			Synaptic CTDs	non-transferred strand of vDNA
Ile-272	I272E	Ile-268	CTD bridge	inter-teramer
			CTD Bridge	from CTD bridge hydrophobic interactions in the same tetramer
Gln-207, Gln-208	QQ207GP		CCD-CTD linker	intra-tetramer
Gln-211, Gln-212, Gln-213	QQQ211PGG		CCD-CTD linker	intra-tetramer

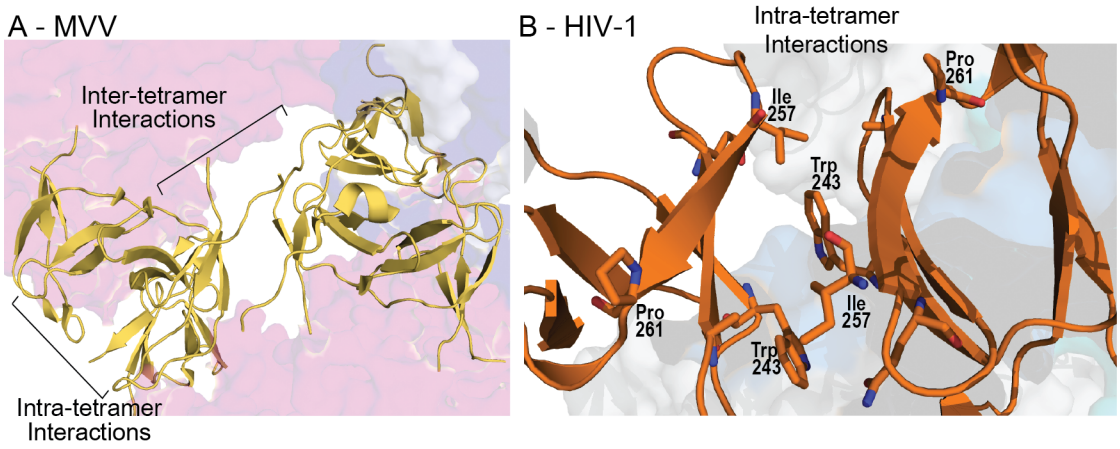

Figure 7-1 CTD-CTD interactions.

Intra-tetramer interactions observed in both (A) MVV (5M0Q) and (B) HIV-1 intasomes. Selected intra-tetramer interactions in HIV-1 intasome are indicated in sticks.

7.1.2 Biochemical analysis of the IN proteins

All the amino acid substituted proteins expressed in bacterial cells, and the majority of these proteins could be purified with sufficient yields and similar elution profiles on heparin affinity chromatography, suggesting no drastic effects on the overall protein folding. One exception was V263E mutant that displayed reduced yields. This MVV IN mutant is structurally analogous to previously analysed V260E HIV-1 IN, shown to be misfolded due to potential disturbances in its intermolecular interactions with other IN domains or IN folding in general (Lutzke and Plasterk, 1998). Although V263E MVV IN was purified with reduced yields, this protein's elution profile from heparin Sepharose was very similar to that of the WT MVV IN. Moreover, MALLS analysis of this mutant showed that it forms a minimally dimeric species. Thus, the V263E MVV IN is unlikely to be grossly misfolded.

As expected, mutations that target CTD-CTD contacts had profound effects on IN multimerization (Figure 3-6). The mutants displayed reduced abilities for tetramerisation and failed to assemble into higher order multimers at the higher input concentrations of 8 mg/ml. In agreement with previously made observations in HIV-1

IN V260E (Kalpana et al., 1999), MVV IN V263E was also highly compromised for multimerisation into higher order oligomers. Oligomerisation of lentiviral INs into at least tetramers is critical for strand transfer activity *in vitro*, and abrogation of HIV-1 IN tetramerisation rendered the virus non-infectious (Hare et al., 2009a, Elliott et al., 2020). Since the mutants reported in this study had reduced abilities for tetramerisation, it is unsurprising to observe that the majority of them were defective for strand transfer activity *in vitro*. Only four of the mutant INs, Y261A, F223A, Y261E, V263E, displayed weakly detectable levels of strand transfer activities. Single amino acid substitutions were previously introduced into the residues Lys-240, Leu-242, Ile-257, Val-259 and Arg-269 of HIV-1 IN, which are part of the CTD-CTD interfaces in the higher order STC assembled and they failed to generate FS integration products *in vitro* (Passos et al., 2017). The MVV IN mutants that were defective for strand transfer activity failed to assemble into intasomes, while the partially active mutants assembled into intasomes with reduced capabilities. These observations indicated that the CTD-CTD interface interactions are of great importance for the overall functionality of the lentiviral intasomes *in vitro*.

MVV IN R231E was generated as an additional control, along with the enzyme active site (E154Q) and tetramerisation defective (H12N) mutants. MVV IN Arg-231 is present in the loop between beta-strands (β1 and β2) in the IN CTD, but within the intasome it is outside the CTD-CTD interface interactions (Figure 3-1). Arg-231 in both HIV-1 and MVV intasomes and STC assemblies interacts with vDNA as well as target DNA (Ballandras-Colas et al., 2017, Passos et al., 2017). Introducing the negatively charged Glu substitution at PFV IN Arg-329 (which is structurally analogous to HIV-1 and MVV IN Arg-231) and HIV-1 IN Arg-231 altered the preferences for target DNA sequences, flanking the integration sites (Maertens et al., 2010, Demeulemeester et al., 2014). Moreover, PFV IN R329E was highly defective for strand transfer activity (Maertens et al., 2010), which is in agreement with the observation made with MVV IN R231E in this study (Figure 3.9).

The α-helical configuration of the CCD-CTD linker was destabilised by substituting the Gln stretches with Pro and Gly residues. Such mutations led to profound loss of strand transfer activity *in vitro*, partly explained as a consequence of these mutants' inability to multimerise and assemble into stable intasomes. Likewise, increasing the

length of the CCD-CTD linker, by replacing the three ^{211}Gln-Gln-Gln^{213} with 11 amino acid residues from the PFV IN CCD-CTD linker (^{304}His-Pro-Ser-Thr-Pro-Pro-Ala-Ser-Ser-Arg-Ser^{314}, known to exist in an extended conformation in the context of the PFV intasome), led to a failure to multimerise beyond dimers in solution and a loss of intasome assembly. The inability of the PFV IN linker to function within the MVV intasome is not surprising. Indeed, within the PFV intasome this stretch of residues is involved in a network of important interactions, including vDNA (Hare et al., 2010). Moreover, His-304 makes contacts with Arg-114 of the NTD-CCD linker, while Arg-313 makes specific contacts with the reactive strand of vDNA (cytosine 11) and Ala-108 of the NTD-CCD linker (Hare et al., 2010). The positioning of the synaptic CTDs in PFV intasome is therefore a consequence of the interactions made by the CCD-CTD linker, with its length aiding this process. Nonetheless, the results allow us to conclude that simply extending the MVV CCD-CTD linker with a stretch of residues that do not form an α-helix is not compatible with functionality *in vitro*. A summary of the observations made with all the mutants is outlined in Table 7-2.

In isolation (produced as a recombinant protein in the absence of vDNA and other viral and host factors), PFV IN is a monomeric protein (Delelis et al., 2008, Valkov et al., 2009, Gupta et al., 2012), which forms tetramers in the context of the intasome (Figure 7-2A). By contrast, MVV and HIV-1 INs are at least tetramers in solution in the absence of vDNA (Wang et al., 2001, Hare et al., 2009a) (Figure 7-2B). Thus, the MVV intasome assembles from individual tetramers as building blocks. The question arises as to why these tetramers or the higher order octamers of lentiviral INs found in solution not sufficient for the intasome assembly without the aid of the additional multimers, and for catalysing the integration reaction. The architecture of both HIV-1 and MVV tetramers formed in solution (Figure 7-2B and C) (Wang et al., 2001, Hare et al., 2009a), is drastically different to the intasomal tetramer assembly observed within the PFV CIC (Figure 7-2A) (Hare et al., 2010). Firstly, the intasome only assembles in the presence of vDNA, as a consequence of the IN subunit rearrangements, including rearrangements observed in the HIV-1 IN CCD-CTD linker (Zhao et al., 2008). Although within MVV IN tetramer there are two enzyme active sites present within proximity to each other, their orientation is not similar to the active sites within the CIC (Figure 7-2D and E). In place of the synaptic CTDs observed in the PFV intasome, NTDs are present in the MVV and HIV-1 IN tetramers

(Figure 7-2D and E). The structure of the HIV-1 intasome was previously modelled based on the PFV intasome structure, as a tetrameric complex (Krishnan et al., 2010). However, the CCD-CTD linkers were assumed to attain an extended conformation devoid of secondary structure, in order to allow the synaptic positioning of the CTDs. This is in fact not the case with MVV and HIV-1 intasome assemblies, where in the CCD-CTD linkers remain in their α-helical confirmations (Figure 7-2F). As reported in this study, the α-helical conformation of the CCD-CTD linker is important for the tetramerisation of IN subunits as well as intasome assembly *in vitro*, that precedes the catalysis of strand transfer.

Table 7-2 Summary of observations made with mutants

	Mutation	Description	Average molecular masses of INs [a]	Strand Transfer Activity[b]	Intasome Assembly[c]	Viral Infectivity[d]	vDNA qunatity, 8 h post infection[e]	
							Early RT products[f]	LRT products[g]
1	WT		+++	100%	+++	100%	100%	100%
2	E154Q	CCD DDE (active site)	+++	0.003%	+++	0.8%	~250%	~100%
3	H12N	NTD HHCC motif	-	0.002%	-	0.9%	~350%	~500%
4	R231E	CTD control	+++	1.000%	++	N/A	N/A	N/A
5	F223A	Inter-tetramer	+	0.600%	-	1.5%		150%
6	V263E	Intra-tetramer	+	0.400%	-	1.0%		140%
7	Y261A	Intra-tetramer	++	3.000%	+	1.5%		65%
8	Y261E	Intra-tetramer	-	1.500%	+	N/A		N/A
9	I272E	Inter-tetramer	+	0.002%	-			
10	Y225A	Inter-tetramer	++	0.004%	-			
11	V252A	Intra-tetramer	++	0.002%	-			
12	V252D	Intra-tetramer	+	0.001%	-			
13	W245E	Intra-tetramer	+	0.004%	-			
14	W245L	Intra-tetramer	++	0.002%	-			
15	QQQ211PGG	CCD-CTD linker	-	2.000%	++	0.9%		130%
16	QQ207GP	CCD-CTD linker	-	0.100%	-	N/A		N/A
17	QQQ211PFV linker	11 amino acid insertion from the long flexible PFV CCD-CTD linker; CCD-CTD linker	-	0.090%	-			

[a] '+++' for ≥160 kDa; '++' for 120 to 159 kDa; '+' for 80 to 119 kDa; '-' for 40 to 79 kDa (Figure 3.6)

[b] % from qPCR, w.r.t WT (Figure 3.10)

[c] '+++' for over 20 mAu; '++' for over 5 mAu; '+' for over 2 mAu; '-' for no peak (Figure 3.12)

[d] % of RLU w.r.t WT (Figure 4.9)

[e] Quantification of vDNA 8 h post infection

[f] Copy number of Luc DNA (% w.r.t WT at 8 h post infection) (Figure 4.10)

[g] LRT quantification (% w.r.t WT at 8 h post infection) (Figure 4.12)

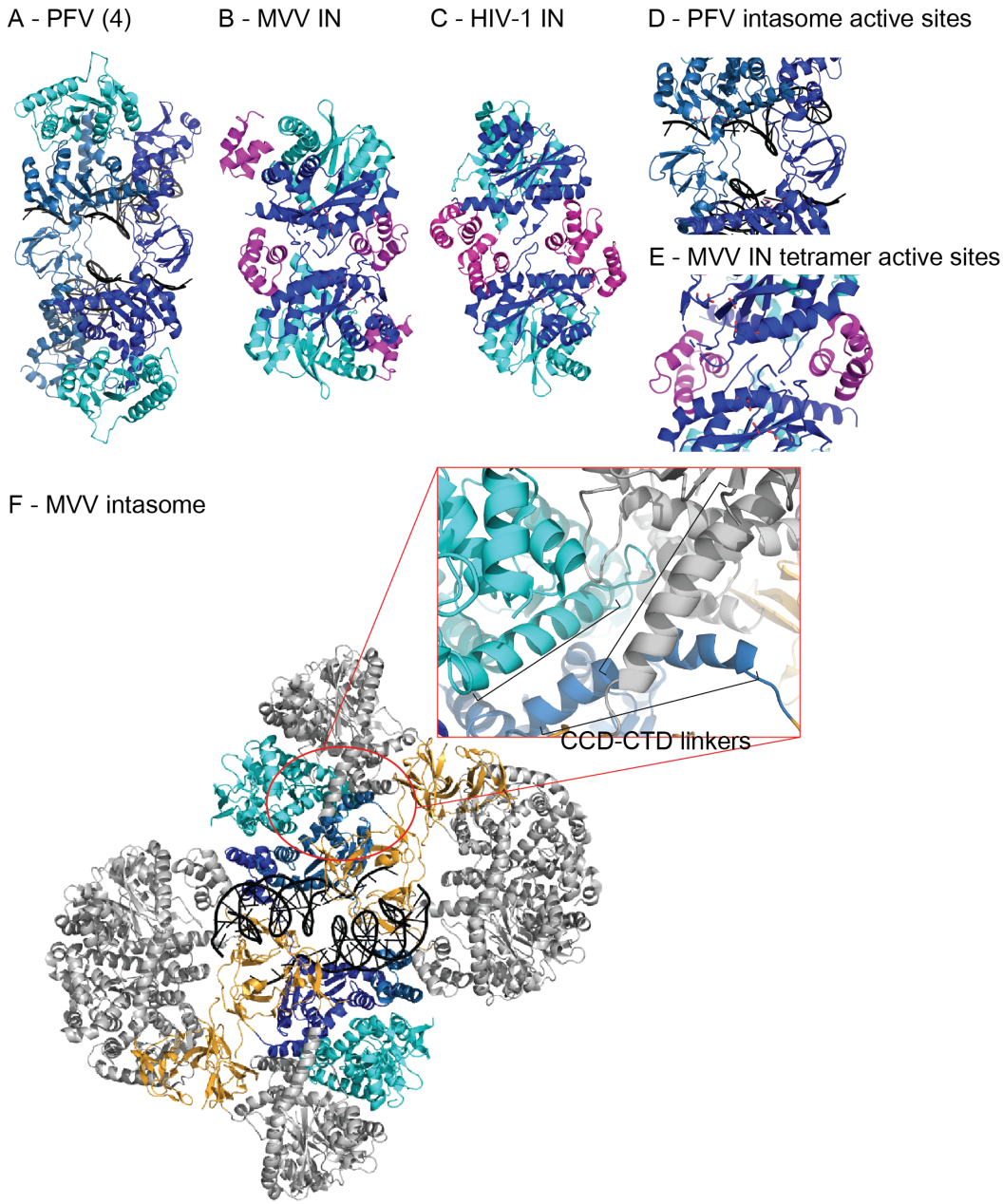

Figure 7-2 Intasome and IN tetramers.

(A) PFV intasome (PDB ID 3OY9); (B) MVV IN (PDB ID 3HPH) and (C) HIV-1 IN (PDB ID 1K6Y) crystal structures of IN NTD-CCD tetramers with CCDs in shades of blue and NTDs in magenta; Enzyme active sites of (D) PFV intasome and (E) MVV IN; (F) MVV intasome (PDB ID 5M0Q), CIC in shades of blue and CTDs in gold, with the CCD-CTD linkers zoomed in and vDNA strands in black.

7.1.3 Importance of the hexadecameric IN assembly for MVV infectivity in a single-cycle vector system

The mutations introduced in the MVV IN to target the IN-IN interactions observed within the intasome are novel and have not been previously characterised biochemically or tested for their infectivity in the context of viral infection. The majority of the mutant MVV IN proteins reported in this study were defective for multimerisation and strand transfer activities *in vitro*. However, the vDNA ends in cells are protected by an extended protein complex (Wei et al., 1997, Chen et al., 1999) and perhaps the intasome is stabilised by the additional interactions within this complex. It was hence important to test the necessity of the hexadecameric intasome assembly for MVV infection, to establish the importance of the unique features for the intasome assembly and viral infectivity.

The mutants introduced into the MVV viral system were carefully thought out. Based on structural superposition of the MVV and HIV-1 IN CTDs, the MVV IN residues Phe-223, Arg-231, Trp-245, Val-252, Val-263, and Tyr-261 are structurally equivalent to HIV-1 IN Phe-223, Arg-231, Trp-243, Val-250, Val-260, and Ile-257, respectively. Mutations of HIV-1 IN Arg-231 and Val-250 were reported to interfere with $tRNA^{Lys3}$ packaging in viral particles, thereby preventing reverse transcription (Phongsavanh et al., 2020). Although R231E was introduced in MVV IN as an additional control for the biochemical analysis, R231E and V252A/D were not selected for experiments with the MVV viral system. HIV-1 IN W243E and V250E were shown to cause class II phenotypes (Wilkinson et al., 2009, Ceccherini-Silberstein et al., 2010), reportedly due to a loss of interaction with RT (Tekeste et al., 2015). Since the mutations introduced at these positions in MVV IN, W245E/L and V252A/D, were inactive for detectable levels of strand transfer activity *in vitro* and because of their potential to cause reverse transcription defects, these mutants were likewise not selected for the infectivity analyses, as they were less likely to produce informative phenotypes. From the set of MVV IN linker mutants, QQQ211PGG was selected for virus assays, because this variant displayed detectable levels of strand transfer activity *in vitro*. HIV-1 IN V260E was shown to be defective specifically for the integration step in viral infection (Kalpana et al., 1999, Elliott et al., 2020). MVV IN F223A and V263E were partially active for strand transfer

activity *in vitro*, hence they were tested in the viral context for infectivity, along with Y261A and QQQ211PGG. Interestingly, all the mutants tested were defective specifically for vDNA integration in the viral infection. Although HIV-1 IN F223A was shown to be defective for vDNA synthesis (Nomaguchi et al., 2014), that was not the case in the case of the equivalent MVV IN F223A mutant.

Currently, there is no evidence for the requirement of the extended oligomeric state of the IN assembly for the pleiotropic roles played by lentiviral IN. Since the minimal units formed by lentiviral INs in solution are tetramers, and the intasome complex is only formed in the presence of vDNA, after reverse transcription, it seems likely that the INs exist as tetramers within the mature virions. Recent data indicate that HIV-1 IN in mature virions is tetrameric (Elliott et al., 2020). Moreover, HIV-1 IN mutants that were unable to form tetramers were also unable to interact with viral genomic RNA resulting in reverse transcription defects (Elliott et al., 2020).

In case of the MVV IN mutants studied, only Y261A, F223A and V263E assembled into tetramers at the highest concentration tested *in vitro*, while QQQ211PGG did not form tetramers in solution (Table 7-2). The tetramerization-defective H12N is well studied in HIV-1 IN and consistently shown to be non-infectious due to defects in virion maturation and reverse transcription (Engelman et al., 1995, Lu et al., 2005c, Elliott et al., 2020). Interestingly, all our mutants, including MVV IN H12N were not defective for vDNA synthesis during infection (Table 7-2). Two distinct reverse transcription intermediates (early and late vDNA products) were measured, and under both these conditions, vDNA synthesis of H12N was higher than that of WT vector (Table 7-2). It can only be assumed that the requirement of IN tetramerisation for viral genome interaction during virion maturation and vDNA synthesis is specific to HIV-1, or the MVV mutants tested were able to assemble into tetramers in the context of the virus. Visualising the mutant MVV particles by TEM may help understand the consequences of the IN mutations introduced during virion morphogenesis.

All the observations made with the mutant INs in this thesis consistently reinforce the importance of the hexadecameric assembly of the lentiviral intasome, even for vDNA integration in the context of the virus. However, these observations do not explain the function of the extended intasome architectures observed in lentiviruses.

7.1.4 Significance of LEDGF/p75 for MVV integration in HEK-293T and CPT cells

Among retroviruses, HIV-1 and other members of the lentiviral genus, are unique in their ability to infect non-dividing cells and for the strong preference for integration into active transcription units (TUs) within gene-dense domains (Schroder et al., 2002, Shun et al., 2007). The targeting of HIV-1 and other primate lentiviral integration to TUs is largely explained by the property of the IN protein – its interaction with LEDGF/p75, while the HIV-1 preference for gene-dense chromosomal domains is explained by the CA-CPSF6 interaction (Sowd et al., 2016, Achuthan et al., 2018, Li et al., 2020). The interaction with LEDGF/p75 and HRP2 is unique to and conserved among lentiviral INs, including MVV IN (Busschots et al., 2005, Cherepanov, 2007), (Table 7-3). In the cases of HIV-1, HIV-2, BIV, EIAV, and FIV, it was demonstrated that this interaction is responsible for the marked preference of these lentiviruses to integrate into TUs (Llano et al., 2004, Ciuffi et al., 2005, Llano et al., 2006, Shun et al., 2007, Schrijvers et al., 2012b, Wang et al., 2012, Li et al., 2020). In the absence of LEDGF/p75, HIV-1 integration targeting is facilitated by HRP2 (Schrijvers et al., 2012b, Wang et al., 2012), whose interaction with IN is markedly weaker (Cherepanov et al., 2004). Both LEDGF/p75 and HRP2 are implicated in DNA DSB repair in mammals (Daugaard et al., 2012, Baude et al., 2016) by homologous end joining. They were also shown to compensate for the FACT (facilitates chromatin transcription) protein complex in differentiated cells and termed FACT-like factors to continue stalled transcription (LeRoy et al., 2019). It is currently unclear if these activities are involved in post-integration DNA repair.

Although it was previously reported that like other lentiviruses, MVV displays preferences for TUs (Ballandras-Colas et al., 2017), it was not known if this property depends on LEDGF/p75. Expansion of the intasomal structure is expected to allow interaction with multiple copies of LEDGF/p75. Indeed, MVV intasome can interact with as many as 16 LEDGF/p75 molecules simultaneously *in vitro* (Ballandras-Colas et al., Manuscript in preparation). Our working hypothesis is that this ability could allow lentiviruses to recognise regions of chromatin with local enrichment of LEDGF/p75 for more selective integration into bearing suitable epigenetic profile.

In the absence of LEDGF/p75, HIV-1 integration was shifted away from gene bodies, towards TSSs and CpG islands and more so in the absence of both LEDGF/p75 and HRP2 in various cell systems tested. (Llano et al., 2004, Ciuffi et al., 2005, Llano et al., 2006, Shun et al., 2007, Schrijvers et al., 2012b, Wang et al., 2012). In addition, a reduction in integration and infectivity was also observed in the absence of these host proteins (Llano et al., 2006, Shun et al., 2007, Schrijvers et al., 2012b, Wang et al., 2012). Similar observations in the absence of LEDGF/p75 were also made with the other lentiviruses, including BIV, EIAV and FIV (Li et al., 2020). Based on the observations made *in vitro*, the necessity of LEDGF/p75 for MVV infectivity and integration site selection was tested in the context of the virus.

Initially, an expected decrease in WT MVV infectivity was observed in human LKO and LHKO cells, similar to the observations made with HIV-1. However, re-expression of LEDGF/p75 in human LHKO cells did not reverse the MVV phenotype. The phenotypes observed in these cells were hence attributed to the clonal nature of the knockout cell lines. Ovine CPT-3 cells were depleted of LEDGF/p75 and HRP2 using CRISPR/Cas9-based methods, to generate both clonal and polyclonal knockout cells. Unlike the infectivity defects observed in human cells, a significant reduction in WT MVV infectivity was not observed in the sheep knockout cell lines. These observations are in line with the interpretations of the phenotype in human LKO and LHKO cells as a clonal effect. It is possible that the LKO clone may express insufficient levels of a host factor required for MVV infectivity.

The unexpected phenotypes in the human LKO and LHKO cells infected with MVV IN E154Q virus, wherein an increase in luciferase expression was observed in comparison with HEK-293T cells, was very surprising. Interestingly, a similar phenotype was also observed in LKO cells infected with the IN active site mutant form (D64N/D116N) of HIV-1 (see Figure 2 in Sowd et al., 2016), even though the reason for it was not addressed or discussed in the original publication. It is highly unlikely that IN active site mutant regains activity in the absence of LEDGF/p75, and indeed the infectivity of E154Q MVV vector in LKO/LHKO cells was not sensitive to inhibition by DTG. With both LEDGF/p75 and HRP2 implicated in homologous recombination during DSB repair (Daugaard et al., 2012, Baude et al., 2016), the observed increase in IN independent insertion of vDNA in the absence of these host

factors, could potentially be a consequence of the non-homologous end joining (NHEJ) DNA repair pathway.

However, an appreciable phenotype was not observed in sheep cells depleted for LEDGF/p75 and HRP2. Sheep is the natural target organism for MVV infection and the lack of a phenotype in LKO and LHKO CPT-3 cells strongly suggests that LEDGF/p75 is not essential for MVV infectivity, unlike the observations made *in vitro*. It is possible that a host factor that might be contributing to PIC stability is lacking in HEK-293T-derived LKO and LHKO cells.

Although a direct effect of LEDGF/p75 and HRP2 depletion on the levels of MVV infectivity were not observed in this study, interesting observations were made with the MVV integration site selections in both human (HEK-293T) and ovine (CPT-3 cells). The information available with HIV-1 integration site sequencing (Sowd et al., 2016, Li et al., 2020) allowed for a seamless comparison of the integration patterns between HIV-1 and MVV in matched cells. In particular, HIV-1 integration is targeted towards active TUs within gene-dense regions of human genome. This phenomenon, is a combined effect of two specific virus-host interactions: (*i*) between lentiviral IN and LEDGF/p75, which directs integration towards bodies of TUs and (*ii*) between capsid and CPSF6, which directs integration towards gene-dense regions and nuclear speckles. Similar to HIV-1, MVV preferentially integrated into TUs, with strong preference for those with higher transcriptional activity. However, by contrast to HIV-1, both in human and sheep cells, MVV did not select for gene-dense regions. In this respect, the distribution of MVV integration sites was very similar to those of other non-primate lentiviruses (EIAV, BIV and FIV) reported in a recent study from Engelman laboratory (Li et al., 2020). This is consistent with the idea that interaction with CPSF6 is a primate specific to primate lentivirus lineage (Li et al., 2020).

While the average gene density at the sites of MVV integration did not significantly change, abrogation of LEDGF/p75 and HRP2 expression greatly reduced the fraction of MVV integration into TUs. The remaining genic integration events displayed significantly reduced correlation with transcriptional activity. Similar to the observations made in HIV-1, MVV integration in the absence of LEDGF/p75 displayed a significantly increased preference for GC-rich regions. This observation is consistent with the presence of the two AT-hook motifs in LEDGF/p75, that help

tether it to AT-rich regions of the chromatin (Turlure et al., 2006). Apart from the increased preference for integration in gene dense regions in HIV-1, all the other lentiviral integration preferences are consistently present throughout the genus.

The only retroviruses known to cause disease in humans (lentiviruses HIV-1 and HIV-2, and deltaretroviruses HTLV-I and HTLV-II), emerged as zoonosis of simian viruses. Both HIV and HTLV target $CD4^+$ T cells in particular, but their clinical manifestations are different. HTLV was prevalent among the indigenous populations, before the emergence of the modern world (Paiva and Casseb, 2015), in comparison, the first HIV-1 group M common ancestor was dated to 1908 (Tebit and Arts, 2011). Compared to HIV-1, HTLV-I evolved a very different infection course, and is thought to primarily replicate in proviral state by clonal proliferation of infected cells (Igakura et al., 2003). Another notable difference among these viruses lies in their integration site preferences. HIV-1 exhibits a strong integration preference within gene bodies in actively transcribing genes (Schroder et al., 2002). Apart from a minor preference for TSSs, no other obvious features were observed with HTLV-I integration site distributions (Derse et al., 2007).

The human genome consists of relatively higher number of pseudogenes than the other genomes analysed (Table 7-3), with 0.3 pseudogenes present for every functional gene. The pseudogenes are generally without any function and considered to be defective copies of other functional genes within the genome (Vanin, 1985). ASLV infects chicken, whose genome is not composed of as many pseudogenes (Table 7-3) and due to its weak preference for actively transcribing genes (Mitchell et al., 2004), its octameric intasome assembly maybe sufficient to facilitate integration into active chromatin. PFV infects chimpanzees, with a tetrameric intasome complex and displays a weak preference for TSSs and CpG islands (Shun et al., 2007). Although HTLV-I is presented with the similar challenges of the human genome, its integration profile is very different to that of HIV-1 and seems to be more similar to the one observed in PFV. While this is highly speculative, it is possible that lentiviruses may have evolved to attain the extended intasome architectures to specifically make simultaneous contacts with multiple copies of LEDGF/p75 and/or HRP2, in order to help them manoeuvre through the rather complicated human

genome and integrate selectively into the chromosomal locations to facilitate a more robust proviral expression.

Table 7-3 Comparison of retroviral intasomes and integration preferences

	Virus	Target organism	ID[a]	Intasome assembly (IN units)	Host factor interactions for targeting integration	Integration preference	Genes [b]	Pseudo-genes [b]	Ratio [b]
1	MVV	Sheep	83	16	LEDGF/p75, HRP2	Gene bodies [c]	33624	5223	1:0.16
2	HIV-1	Human	51	~12	LEDGF/p75, HRP2	Gene bodies [d]	61238	18367	1:0.3
3	HTLV-1	Human	51	4	PP2A	TSSs and CpG islands [e]			
4	PFV	Chimpanzee	202	4	N/A	TSSs and CpG islands [d]	40986	7595	1:0.19
5	RSV	Chicken	111	8	N/A	weak preference for active genes [f]	24407	262	1:0.01

[a] NCBI genome ID for sheep (*Ovis aeres*), human (*Homo sapiens*), chimpanzee (*Pan troglodytes*) and chicken (*Gallus gallus*)
[b] Number of genes and pseudogenes and the ratio of genes:pseudogenes
[c] This study
[d] Schroder et al., 2002, Shun et al., 2007
[e] Derse et al., 2007
[f] Mitchell et al., 2004

7.2 Future directions

7.2.1 Investigating HIV-1 IN mutants

IN CTDs in the MVV intasome play a vital role in its function and interfering with them caused drastic loss of function defects. A similar observation was made by introducing amino acid substitutions in the α-helical CCD-CTD linker. Amino acid substitutions were not introduced in the IN NTD in this study. The NTDs from the intasomal MVV IN chains A and I, the only two subunits that provide active sites for catalysis, are part of the catalytic tetramers (Figure 1-11) interact with their CCDs *in trans* in the intasome, as observed in PFV intasome (Figure 1-6). NTDs also play a role in tetramerisation of the lentiviral IN subunits (Hare et al., 2009a). Based on the observations reported in this thesis and the overall roles played by the IN domains in ensuring the functionality of the protein, introducing more mutations to disrupt the intasome interactions may only result in total loss of function. However, some of the mutations in the sequentially conserved residues, such as F223A can be introduced into HIV-1 IN to be analysed biochemically for IN strand transfer activity and oligomerisation in solution, while testing their infectivity profiles. This may aid in further understanding the necessity of the extended lentiviral intasome assembly.

7.2.2 Presence of other host binding partners

CA-CPSF6 interaction was observed only in primate lentiviruses (Li et al., 2020), and this interaction aids is guiding PICs towards gene-dense regions of the genome, while LEDGF/p75 plays a more local role in integration site selection, directing integration into active TUs (Sowd et al., 2016, Achuthan et al., 2018). The subtle differences in targeting integration towards highly expressive genes in human and ovine cells infected with MVV viruses, suggests the presence of potential ovine specific host factors. These host factors may either interact with MVV CA, similar to the HIV-1 CA-CPSF6 interaction or in a previously unknown mechanism of interaction. It will be of interest to identify these hypothetical host factor(s) and establish the roles they would play in non-primate lentiviral infection. Indeed, the MVV infectivity defects observed in human LKO and LHKO cells that could not be reversed by re-expression of LEDGF/p75 seem to point in this direction.

7.3 Final conclusions

The main question at the start of this project was whether it is possible to assemble a functional MVV intasome *in vitro* from reduced number of IN subunits, instead of the observed quadruple IN tetramer complex (Ballandras-Colas et al., 2017) and to check its compatibility for viral infection. It was revealed that the hexadecameric intasome assembly is important for lentiviral integration and this extended architecture is relevant in the context of the virus. Although its functional consequences will need extensive investigation, its interaction with LEDGF/p75 is important to aid integration into TUs. Insertion of vDNA ends into optimal regions of the host chromosomes by lentiviral integration machinery is a highly regulated process, not only by the interaction of the intasome with specific host factors, but also the host factor interactions made by other viral elements, for example the primate lentiviral CA-CPSF6 interaction. The understanding of this mechanism, taking into account all the pieces of the puzzle, will potentially aid in a better understanding of how and why lentiviruses evolved to attain the extended intasome architecture.

Chapter 8. Appendix

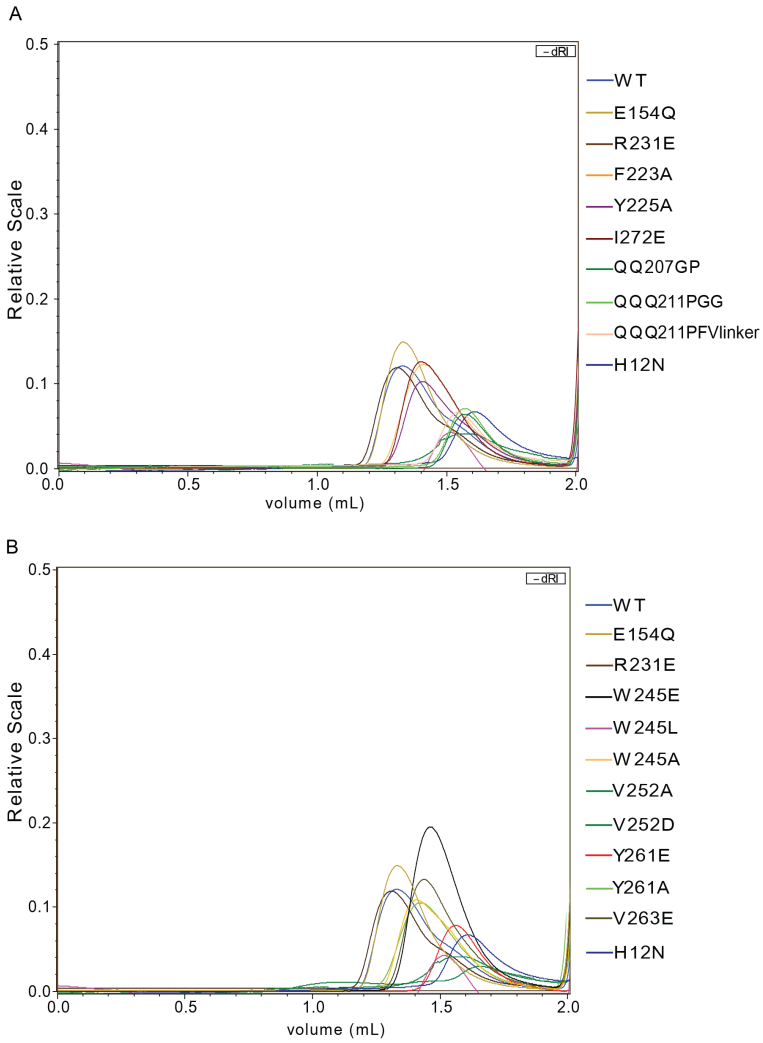

Figure 8-1 Chromatograms from SEC-MALLS of mutants INs at 1 mg/ml

Chromatograms from WT and mutant IN subunits injected at 1 mg/ml into Superdex 200 Increase 3.2/300 column. Elution profiles of IN WT, E154Q, H12N and R231E with (A) Inter-tetramer mutants – F223A, Y225A and I272E, CCD-CTD linker mutants – QQ207GP, QQQ211PGG and QQQ211PFVlinker, (B) Intra-tetramer mutants – W245E, W245L, W245A, V252A, V252D, Y261E, Y261A and V263E.

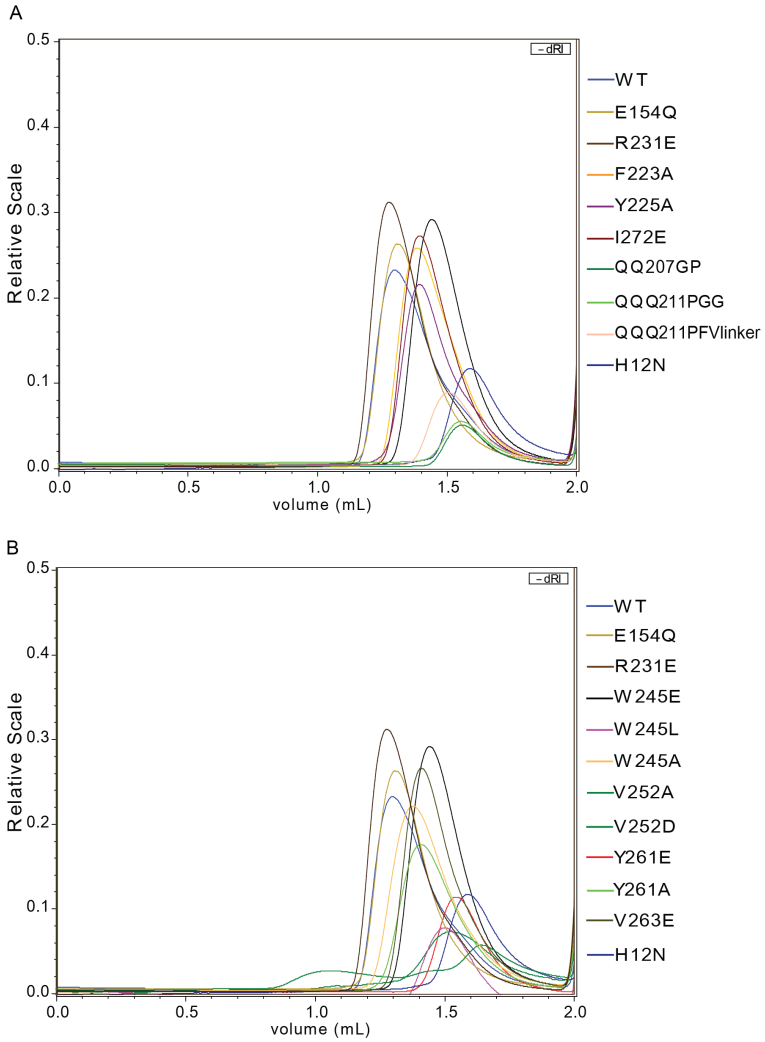

Figure 8-2 Chromatogram from SEC-MALLS of mutant INs at 2 mg/ml.

Chromatograms from WT and mutant IN subunits injected at 2 mg/ml into Superdex 200 Increase 3.2/300 column. Elution profiles of IN WT, E154Q, H12N and R231E with (A) Inter-tetramer mutants – F223A, Y225A and I272E, CCD-CTD linker mutants – QQ207GP, QQQ211PGG and QQQ211PFVlinker, (B) Intra-tetramer mutants – W245E, W245L, W245A, V252A, V252D, Y261E, Y261A and V263E.

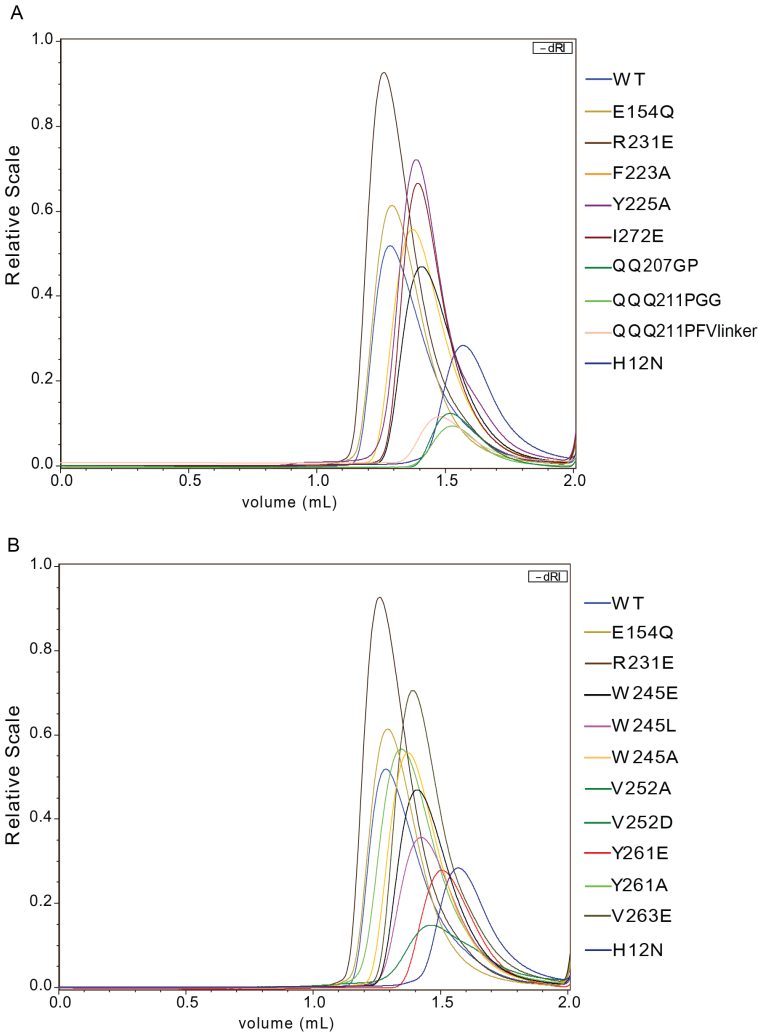

Figure 8-3 Chromatograms from SEC-MALLS of mutant INs at 8 mg/ml.

Chromatograms from WT and mutant IN subunits injected at 8 mg/ml into Superdex 200 Increase 3.2/300 column. Elution profiles of IN WT, E154Q, H12N and R231E with (A) Inter-tetramer mutants – F223A, Y225A and I272E, CCD-CTD linker mutants – QQ207GP, QQQ211PGG and QQQ211PFVlinker, (B) Intra-tetramer mutants – W245E, W245L, W245A, V252A, V252D, Y261E, Y261A and V263E.

```
SHEEP       MTRDFKPGDLIFAKMKGYPHWPARVDEVPDGAVKPPTNKLPIFFFGTHETAFLGPKDIFP    60
Human       MTRDFKPGDLIFAKMKGYPHWPARVDEVPDGAVKPPTNKLPIFFFGTHETAFLGPKDIFP    60
            ************************************************************

SHEEP       YSENKEKYGKPNKRKGFNEGLWEIDNNPKVKFSSQQASAKQSNASSDVEVEEKETSVSKE    120
Human       YSENKEKYGKPNKRKGFNEGLWEIDNNPKVKFSSQQAATKQSNASSDVEVEEKETSVSKE    120
            ************************************:::*********************

SHEEP       DTDHEEKASNEDVTKAIDITTPKAARRGRKRKAEKQVETEEAAVVTTATASANLKVSPKR    180
Human       DTDHEEKASNEDVTKAVDITTPKAARRGRKRKAEKQVETEEAGVVTTATASVNLKVSPKR    180
            ****************:*************************.********.********

SHEEP       GRPAATEVKIPKPRGRPKMVKQPCPSESDMITEEDKSKKKGQEEKQPKKQLKKDEEGQKE    240
Human       GRPAATEVKIPKPRGRPKMVKQPCPSESDIITEEDKSKKKGQEEKQPKKQPKKDEEGQKE    240
            ****************************:******************** *********

SHEEP       EEKPRKEPDKKEGKKEVESKRKNLAKTGVTSTSDSEEEGDDQEGEKKRKGGRNFQTAHRR    300
Human       EDKPRKEPDKKEGKKEVESKRKNLAKTGVTSTSDSEEEGDDQEGEKKRKGGRNFQTAHRR    300
            *:************************************************************

SHEEP       NMLKGQHEKEVADRKRKQEEQMETEQQNKDEGKKPEVKKVEKKRETSMDSRLQRIHAEIK    360
Human       NMLKGQHEKEAADRKRKQEEQMETEQQNKDEGKKPEVKKVEKKRETSMDSRLQRIHAEIK    360
            **********.*************************************************

SHEEP       NSLKIDNLDVNRCIEALDELASLQVTMQQAQKHTEMITTLKKIRRFKVSQVIMEKSTMLY    420
Human       NSLKIDNLDVNRCIEALDELASLQVTMQQAQKHTEMITTLKKIRRFKVSQVIMEKSTMLY    420
            ************************************************************

SHEEP       NKFKNMFLVGEGDSVITQVLNKSLAEQRQHEEANKTKDQGKKGPNKKLEKEQTGSKTLNG    480
Human       NKFKNMFLVGEGDSVITQVLNKSLAEQRQHEEANKTKDQGKKGPNKKLEKEQTGSKTLNG    480
            ************************************************************

SHEEP       GSDAQDSNQPQHNGDSNEESKDNHEASSKKKPSSEERETEISLKDSTLDN    530
Human       GSDAQDGNQPQHNGESNEDSKDNHEASTKKKPSSEERETEISLKDSTLDN    530
            ******.*******:***:********:**********************
```

Figure 8-4 Amino acid sequence alignment of LEDGF/p75.

Amino acid sequence alignment of sheep and human LEDGF/p75, with the IBD sequences in red. Sequences aligned using Clustal Omega Multiple Sequence Alignment Tool (https://www.ebi.ac.uk/Tools/msa/clustalo/)

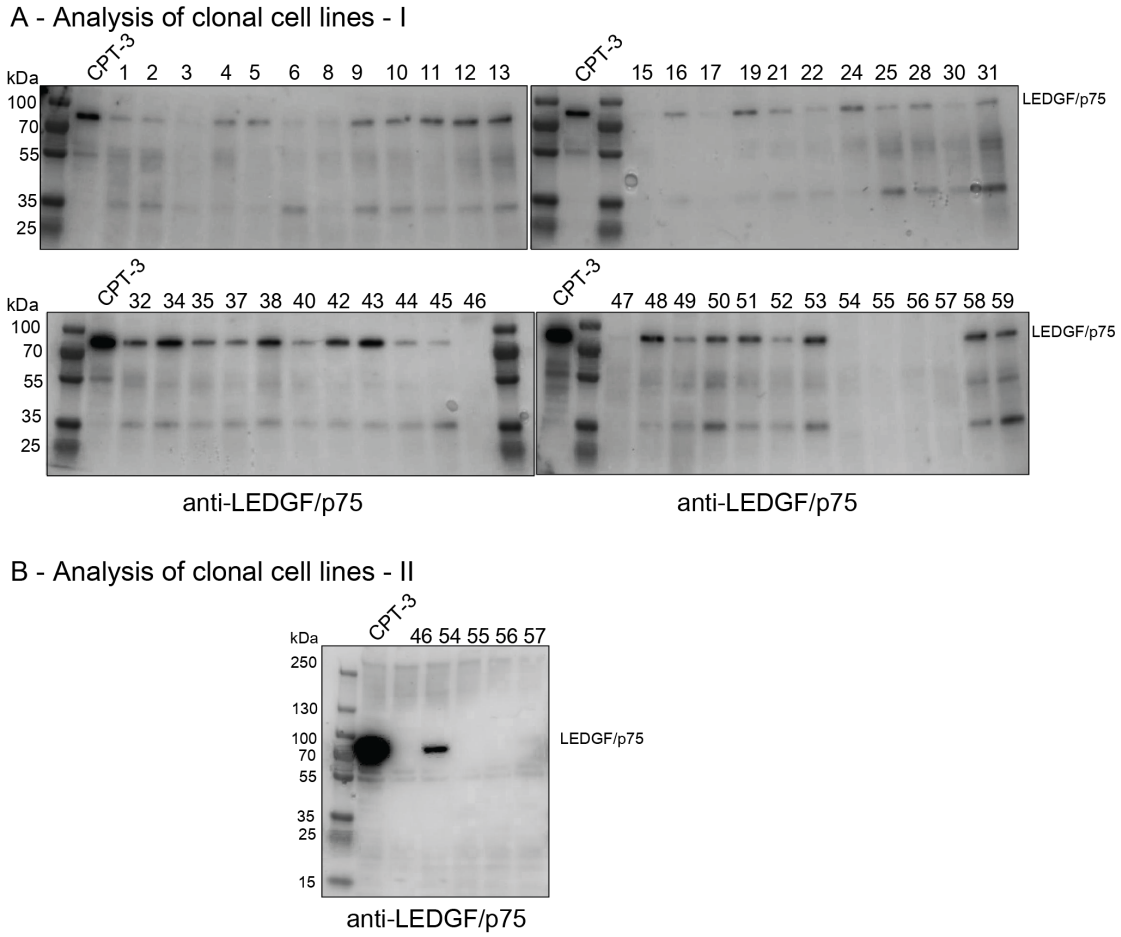

Figure 8-5 Initial analysis of sheep clonal cell lines grown out after transfection with pSpCas9(BB)-2A-GFP-gRNA1/2/3 (*PSIP1*-targeting construct).

Clonal cell lines were tested for expression of LEDGF/p75 by Western blotting (chemiluminescent detection). (A) Initial analysis of clonal cells. gRNA 1 (1 to 29), gRNA 2 (30 to 44) and gRNA 3 (45 to 59). (B) Selected cell lines were re-tested after two more passages.

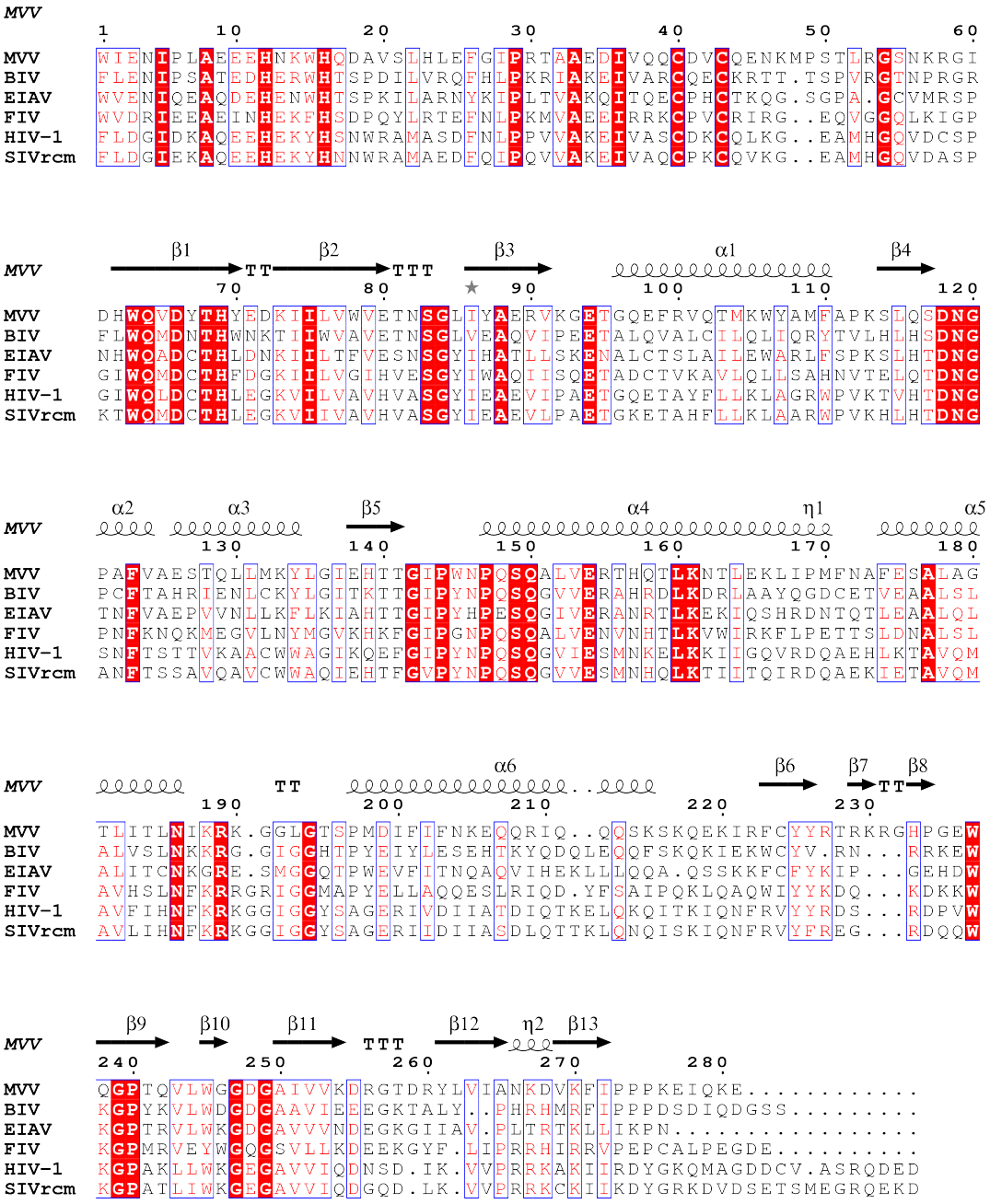

Figure 8-6 Sequence alignment of lentiviral INs.

Amino acid sequences of lentiviral INs aligned in Clustal omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and analysed on ESPript (https://espript.ibcp.fr,).

Table 8-1 Plasmids used in this study

No	Plasmid Name	Use/Description	Restriction sites used	Reference
1	pCPH6P-MVV-IN-WT	Purification of recombinant WT MVV IN	*Xma*I and *Xho*I	Cherepanov, 2007
2	pCPH6P-MVV-IN-E154Q	Purification of recombinant E154Q MVV IN	*Xma*I and *Xho*I	This study
3	pCPH6P-MVV-IN-H12N	Purification of recombinant H12N MVV IN	*Xma*I and *Xho*I	This study
4	pCPH6P-MVV-IN-R231E	Purification of recombinant R231E MVV IN	*Xma*I and *Xho*I	This study
5	pCPH6P-MVV-IN-F223A	Purification of recombinant F223A MVV IN	*Xma*I and *Xho*I	This study
6	pCPH6P-MVV-IN-V263E	Purification of recombinant V263E MVV IN	*Xma*I and *Xho*I	This study
7	pCPH6P-MVV-IN-Y261A	Purification of recombinant Y261A MVV IN	*Xma*I and *Xho*I	This study
8	pCPH6P-MVV-IN-Y261E	Purification of recombinant Y261E MVV IN	*Xma*I and *Xho*I	This study
9	pCPH6P-MVV-IN-I272E	Purification of recombinant I272E MVV IN	*Xma*I and *Xho*I	This study
10	pCPH6P-MVV-IN-Y225A	Purification of recombinant Y225A MVV IN	*Xma*I and *Xho*I	This study
11	pCPH6P-MVV-IN-V252A	Purification of recombinant V252A MVV IN	*Xma*I and *Xho*I	This study
12	pCPH6P-MVV-IN-V252D	Purification of recombinant V252D MVV IN	*Xma*I and *Xho*I	This study
13	pCPH6P-MVV-IN-W245E	Purification of recombinant W245E MVV IN	*Xma*I and *Xho*I	This study
14	pCPH6P-MVV-IN-W245L	Purification of recombinant W245L MVV IN	*Xma*I and *Xho*I	This study
15	pCPH6P-MVV-IN-QQQ211PGG	Purification of recombinant QQQ211PGG MVV IN	*Xma*I and *Xho*I	This study

No	Plasmid Name	Use/Description	Restriction sites used	Reference
16	pCPH6P-MVV-IN-QQ207GP	Purification of recombinant QQ207GP MVV IN	*Xma*I and *Xho*I	This study
17	pCPH6P-MVV-IN-QQQ211PFVlinker	Purification of recombinant QQQ211PFVlinker MVV IN	*Xma*I and *Xho*I	This study
18	pCPH6P-LEDGF/p75	Purification of recombinant LEDGF/p75		Cherepanov, 2007
19	pBluescript SK(-)	Cloning vector, used to modify MVV IN	*Bam*HI and *Xho*I	Dr. Rebecca Newman
20	pCAG-MV-Gagpol-CTEx2	MVV single cycle plasmid encoding WT-MV1 form of Gagpol	*Pas*I and *Dra*III	Dr. David Griffiths
21	pCAG-MV-GagPol-CTEx2-IN.KV1772-*Age*I/*Xho*I	MVV single cycle plasmid encoding WT-KV1 IN, flakned by *Age*I and *Xho*I in Gagpol	*Age*I and *Xho*I	This study
22	pCAG-MV-GagPol-CTEx2-IN.KV1772-*Age*I/*Xho*I-E154Q	MVV single cycle plasmid encoding E154Q form of KV1 IN, flakned by *Age*I and *Xho*I in Gagpol	*Age*I and *Xho*I	This study
23	pCAG-MV-GagPol-CTEx2-IN.KV1772-*Age*I/*Xho*I-H12N	MVV single cycle plasmid encoding H12N form of KV1 IN, flakned by *Age*I and *Xho*I in Gagpol	*Age*I and *Xho*I	This study
24	pCAG-MV-GagPol-CTEx2-IN.KV1772-*Age*I/*Xho*I-F223A	MVV single cycle plasmid encoding F223A form of KV1 IN, flakned by *Age*I and *Xho*I in Gagpol	*Age*I and *Xho*I	This study
25	pCAG-MV-GagPol-CTEx2-IN.KV1772-*Age*I/*Xho*I-Y261A	MVV single cycle plasmid encoding Y261A form of KV1 IN, flakned by *Age*I and *Xho*I in Gagpol	*Age*I and *Xho*I	This study
26	pCAG-MV-GagPol-CTEx2-IN.KV1772-*Age*I/*Xho*I-V263E	MVV single cycle plasmid encoding V263E form of KV1 IN, flakned by *Age*I and *Xho*I in Gagpol	*Age*I and *Xho*I	This study

No	Plasmid Name	Use/Description	Restriction sites used	Reference
27	pCAG-MV-GagPol-CTEx2-IN.KV1772-*Age*I/*Xho*I-QQQ211PGG	MVV single cycle plasmid encoding QQQ211PGG form of KV1 IN, flakned by *Age*I and *Xho*I in Gagpol	*Age*I and *Xho*I	This study
28	pCAG-MV-GagPol-CTEx2-IN$^{D66A/E154A}$	MVV single cycle plasmid encoding WT-MV1 form of Gagpol, with IN mutations-D66A and E154A		Dr. David Griffiths
29	pCVW-CG-GFP	MVV single cycle vector plasmid with CMV-GFP-WPRE expression cassette		Dr. David Griffiths
30	pCMV-MVV-Rev	MVV single cycle plasmid encoding codon-optimised Visna Rev		Dr. David Griffiths
31	pCVW-CG-Luc	MVV single cycle vector plasmid with CMV-Luc-WPRE expression cassette		Dr. Wen Li, Engelman lab
32	pMD2.G	Plasmid encoding VSV-G		Available from Addgene
33	pCG-GagPol	MLV single cycle plasmid encoding WT form of Gagpol		Ulm et al., 2007
34	pCG-VSVG	Plasmid encoding VSV-G		Ulm et al., 2007
35	pQCXIP	Plasmid encoding puromycin resistance gene, for selecting stably transduced cells	*Not*I and *Bam*HI	Clontech
36	pQCXIP-WT	Encoding WT form of *Oa*LEDGF/p75	*Not*I and *Bam*HI	This study

No	Plasmid Name	Use/Description	Restriction sites used	Reference
37	pQCXIP-D366N	Encoding D366N form of *Oa* LEDGF/p75	*Not*I and *Bam*HI	This study
38	pCR4-TOPO-MVV-LRT	Plasmid encoding MVV LRT used to generate standard curve in qPCR reactions to quantify LRT products		This study
39	pSpCas9(BB)-2A-GFP	Plasmid encoding the necessary components for the action of CRISPR/Cas9, with a GFP reporter	*Bbs*I	Ran et al., 2013
40	pSpCas9(BB)-2A-GFP-sgRNA1	Plasmid encoding the necessary components for the action of CRISPR/Cas9 and encoding sgRNA1	*Bbs*I	This study
41	pSpCas9(BB)-2A-GFP-sgRNA2	Plasmid encoding the necessary components for the action of CRISPR/Cas9 and encoding sgRNA2	*Bbs*I	This study
42	pSpCas9(BB)-2A-GFP-sgRNA3	Plasmid encoding the necessary components for the action of CRISPR/Cas9 and encoding sgRNA3	*Bbs*I	This study

Table 8-2 Cell lines used in this study

No	Cell line name	Description	Reference
1	HEK-293T	Parental HEK-293T cells	Prof. Eric M Poeschla
2	HEK-293T LKO	Clonal HEK-293T cells knocked out of LEDGF/p75	Fadel et al., 2014
3	HEK-293T HKO	Clonal HEK-293T cells knocked out of HRP2	Prof. Eric M Poeschla
4	HEK-293T LHKO	Clonal HEK-293T cells knocked out of LEDGF/p75 and HRP2	Prof. Eric M Poeschla
5	CPT-3	Ovine choroid plexus cells, clone-3	Arnaud et al., 2010
6	CPT-7	Ovine choroid plexus cells, clone-7	Arnaud et al., 2010
7	CPT-8	Ovine choroid plexus cells, clone-8	Arnaud et al., 2010
8	CPT-3 LKO 1	Polyclonal CPT-3 cells, knocked out of LEDGF/p75	This study
9	CPT-3 LKO 2	Polyclonal CPT-3 cells, knocked out of LEDGF/p75	This study
10	CPT-3 LKO 3	Polyclonal CPT-3 cells, knocked out of LEDGF/p75	This study
11	CPT-3 LKO 4	Polyclonal CPT-3 cells, knocked out of LEDGF/p75	This study
12	CPT-3 LHKO 1	Polyclonal CPT-3 cells, knocked out of LEDGF/p75 and HRP2	This study
13	CPT-3 LHKO 2	Polyclonal CPT-3 cells, knocked out of LEDGF/p75 and HRP2	This study
14	CPT-3 46	Clonal CPT-3 cells knocked out of LEDGF/p75	This study
15	CPT-3 55	Clonal CPT-3 cells knocked out of LEDGF/p75	This study
16	CPT-3 56	Clonal CPT-3 cells knocked out of LEDGF/p75	This study

Table 8-3 Primers used in this study

	Primer code	Primer sequence (5' to 3')	Description	Reference
1	PC1095	GAGC-CCCGGG-TGGATAGAAAATATTCCCCTAGCAGAAG	Visna MVV IN sense XmaI	Cherepanov, 2007
2	PC1096	GCGC-CTCGAG-TTATTCTTTTTGTATTTCTTTTGGTGG	Visna MVV IN antisense stop/XhoI	Cherepanov, 2007
3	PC1192	CTGGAGATAGCTAACAAAGATGTTAAGTTCATACC	MVV IN V263E sense	This study
4	PC1193	CATCTTTGTTAGCTATCTCCAGATATCTATCTGTGC	MVV IN V263E antisense	This study
5	PC1194	AACACAGGTACTTGCAGGCGGGGACGGTGCGATTG	MVV IN W245A sense	This study
6	PC1195	ACCGTCCCCGCCTGCAAGTACCTGTGTTGGTCCTTGC	MVV IN W245A antisense	This study
7	PC1196	CCAACACAGGTACTTGAGGGCGGGGACGGTGCGATTG	MVV IN W245E sense	This study
8	PC1197	CGCACCGTCCCCGCCCTCAAGTACCTGTGTTGGTCC	MVV IN W245E antisense	This study
9	PC1198	GAAAAAATTCGAGCATGTTATTACAGAACAAGAAAAAGAGG	MVV IN F223A sense	This study
10	PC1199	CTGTAATAACATGCTCGAATTTTTTCTTGTTTTGATTTAC	MVV IN F223A antisense	This study
11	PC1200	CGATTTTGTGCGTACAGAACAAGAAAAAGAGGGCATC	MVV IN Y225A sense	This study
12	PC1201	CTTGTTCTGTACGCACAAAATCGAATTTTTTCTTGTTTTG	MVV IN Y225A antisense	This study
13	PC1202	GATGTTAAGTTCGAGCCGCCACCAAAAGAAATACAA	MVV IN I272E sense	This study
14	PC1203	GGCGGCTCGAACTTAACATCTTTGTTAGCTATCACC	MVV IN I272E antisense	This study
15	PC1204	GAACAAGAAAAGAGGGGCATCCAGGAGAGTGGCAAGG	MVV IN R231E sense	This study

	Primer code	Primer sequence (5' to 3')	Description	Reference
16	PC1205	GATGCCCCTCTTTTCTTGTTCTGTAATAACAAAATCG	MVV IN R231E antisense	This study
17	PC1206	CAGAGGCACAGATAGAGCTCTGGTGATAGCTAAC	MVV IN Y261A sense	This study
18	PC1207	GTTAGCTATCACCAGAGCTCTATCTGTGCCTCTG	MVV IN Y261A antisense	This study
19	PC1208	CCCCTAGCAGAAGAAGAGAATAACAAATGGCATCAAG	MVV IN H12N sense	This study
20	PC1209	CTTGATGCCATTTGTTATTCTCTTCTTCTGCTAGGGG	MVV IN H12N antisense	This study
21	PC1210	CAATCTCAAGCATTAGTACAGAGAACACATCAGACG	MVV IN E154Q sense	This study
22	PC1211	CGTCTGATGTGTTCTCTGTACTAATGCTTGAGATTG	MVV IN E154Q antisense	This study
23	PC1238	CAACACAGGTACTTTTGGGCGGGGACGGTG	MVV IN W245L sense	This study
24	PC1239	CACCGTCCCCGCCCAAAAGTACCTGTGTTG	MVV IN W245L antisense	This study
25	PC1240	GACAGAGGCACAGATAGAGAACTGGTGATAGCTAACAAAG	MVV IN Y261E sense	This study
26	PC1241	CTTTGTTAGCTATCACCAGTTCTCTATCTGTGCCTCTGTC	MVV IN Y261E antisense	This study
27	PC1242	GACGGTGCGATTGCAGTGAAAGACAGAG	MVV IN V252A sense	This study
28	PC1243	CTCTGTCTTTCACTGCAATCGCACCGTC	MVV IN V252A antisense	This study
29	PC1244	GGACGGTGCGATTGACGTGAAAGACAGAGG	MVV IN V252D sense	This study
30	PC1245	CCTCTGTCTTTCACGTCAATCGCACCGTCC	MVV IN V252D antisense	This study

	Primer code	Primer sequence (5' to 3')	Description	Reference
31	PC1302	AAGGAAGGACCAAGAATACAGCAACAAA GTAAATC	MVV IN Q207G/Q208P sense	This study
32	PC1303	GTATTCTTGGTCCTTCCTTATTAAATATAA ATATATCC	MVV IN Q207G/Q208P antisense	This study
33	PC1304	AAAGAATACCGGGACCAAGTAAATCAAAA CAAG	MVV IN Q211P/Q212G/Q213 P sense	This study
34	PC1305	GATTTACTTGGTCCCGGTATTCTTTGTTGT TCC	MVV IN Q211P/Q212G/Q213 P antisense	This study
35	PC1306	AAAGAATACCGGGAGGGAGTAAATCAAAA CAAG	MVV IN Q211P/Q212G/Q213 G sense	This study
36	PC1307	GATTTACTCCCTCCCGGTATTCTTTGTTGT TCC	MVV IN Q211P/Q212G/Q213 G antisense	This study
37	PC1308	GGGTGGATGGATGGGATATTCTTTGTTGTT CCT	PFV CCD-CTD linker antisense 1	This study
38	PC1309	CCCGGAACGAGAGGAGGCTGGAGGGGT GGATGGATGGGATA	PFV CCD-CTD linker antisense 2	This study
39	PC1310	AGCCTCCTCTCGTTCCGGGAGTAAATCAA AACAAGAAAA	PFV CCD-CTD linker sense	This study
40	EV272	CCGTGCAACACCGGAGCGGATCTCGCA	MVV U5 29* upper strand, Visna U5 mutant donor	Ballandras-Colas et al., 2017
41	EV273	GCTGCGAGATCCGCTCCGGTGTTGCACG G	MVV U5 29* bottom strand, Visna U5 mutant donor	Ballandras-Colas et al., 2017
42	EV272-5'-Cy5	5'Cy5-CCGTGCAACACCGGAGCGGATCTCGCA	MVV U5 29* upper strand, Visna U5 mutant donor	This study
43	PC1235	CCGGCTTTCCCCGTCAAGCT	pGEM target qPCR primer 1	This study
44	PC1237	ACACCGGAGCGGATCTCG	MVV donor sense qPCR primer	This study
45	PC1246	CCGGCGGATCCACGAAGAGGTCATTATA CCCCAGGGGAGG	MVV POL sense (BamHI)	This study

	Primer code	Primer sequence (5' to 3')	Description	Reference
46	PC1247	CCGGCCTCGAGACCGTCTATCAGGGCGATGGCCCACTACG	MVV POL antisense (XhoI)	This study
47	PC1252	GGATTTGGATCAACCGGTGCATATTGGATTG	AgeI, sense; upstream IN, pCAG-MV-Gagpol-CTE2X	This study
48	PC1253	CAATCCAATATGCACCGGTTGATCCAAATCC	AgeI, antisense; upstream IN, pCAG-MV-Gagpol-CTE2X	This study
49	PC1254	GAGCAAAAATAGGCTCGAGATAGGGCCTCG	XhoI, sense; downstream IN, pCAG-MV-Gagpol-CTE2X option 1	This study
50	PC1255	CGAGGCCCTATCTCGAGCCTATTTTTGCTC	XhoI, antisense; downstream IN, pCAG-MV-Gagpol-CTE2X	This study
51	PC1263	GGATCAACCGGTGCATATTGGATAGAAAATATTCCCC	AgeI, sense; upstream IN, pCPH6P_MVV_IN	This study
52	PC1335	TCCTGCTCAACTTCCTGTCGAG	F' Roche MS2 RNA (10165948001)	Vermeire et al., 2012
53	PC1336	CACAGGTCAAACCTCCTAGGAATG	R' Roche MS2 RNA (10165948001)	Vermeire et al., 2012
54	PC1566	6[FAM]CGAGACGCTACCATGGCTATCGCTGTAG[TAMRA]	Real time PCR MS2 TaqMan Probe	Li et al., 2015
55	PC1488	TTCCATCTGCCAGGTATCAGG	Real time PCR Luc F' primer TaqMan	Dr. Wen Li, Engelman lab
56	PC1489	CCGGTATCCAGATCCACAAC	Real time PCR Luc R' primer TaqMan	Dr. Wen Li, Engelman lab
57	PC1490	[FAM]TGGGCTCACTGAGACTACATCAGCTATTCTGA[TAMRA]	Real time PCR Luc TaqMan Probe	Dr. Wen Li, Engelman lab
58	PC1599	CACCGCAAGTTTAAGAACATGTTTT	gRNA1_F'_20nt 'sgRNA-1'	This study

	Primer code	Primer sequence (5' to 3')	Description	Reference
59	PC1600	AAACAAAACATGTTCTTAAACTTGC	gRNA1_R'_20nt 'sgRNA-1'	This study
60	PC1601	CACCGCAATGGATTCTCGACTTCAA	gRNA2_F'_20nt 'sgRNA-2'	This study
61	PC1602	AAACTTGAAGTCGAGAATCCATTGC	gRNA2_R'_20nt 'sgRNA-2'	This study
62	PC1603	CACCGCAAAGTTAGTCAGGTTATCA	gRNA6_F'_20nt 'sgRNA-3'	This study
63	PC1604	AAACTGATAACCTGACTAACTTTGC	gRNA6_R'_20nt 'sgRNA-3'	This study
64	PC1605	GAGGGCCTATTTCCCATGATTCC	F' Seq. primer for gRNA on pSpCas9(BB)-2A-GFP	Ran et al., 2013
65	PC1796	[FAM]CGTGGGGCTCGACAAAGAATC[TAMRA]	Real time PCR MVV LRT TaqMan Probe	Dr. Wen Li, Engelman lab
66	PC1803	GGATCCTGGACTGTCAGGACAGAGAAC	Outer F' with BamHI upstream MVV LRT Standard for qPCR	This study
67	PC1804	TCCGGTGTTGCACGGAATTAG	Inner R' MVV LRT Standard for qPCR	This study
68	PC1805	CTGATATCTGAGCTTGCCTGG	Inner F' MVV LRT Standard for qPCR	This study
69	PC1806	CTCGAGTTCTCCTTTGAGCCTTGCTTC	Outer R' with XhoI upstream	This study
70	PC1810	GATGGTTAAGTCATAACCGCAGATG	F' MVV LRT primer for qPCR	This study
71	PC1812	GGTGTCTCTCTTACCTTACTTCAGG	R' MVV LRT primer for qPCR	This study

Reference List

ACHUTHAN, V., PERREIRA, J. M., SOWD, G. A., PURAY-CHAVEZ, M., MCDOUGALL, W. M., PAULUCCI-HOLTHAUZEN, A., WU, X., FADEL, H. J., POESCHLA, E. M., MULTANI, A. S., HUGHES, S. H., SARAFIANOS, S. G., BRASS, A. L. & ENGELMAN, A. N. 2018. Capsid-CPSF6 Interaction Licenses Nuclear HIV-1 Trafficking to Sites of Viral DNA Integration. *Cell Host & Microbe,* 24**,** 392-404.e8.

ADAMS, D. S., KLEVJER-ANDERSON, P., CARLSON, J. L., MCGUIRE, T. C. & GORHAM, J. R. 1983. Transmission and control of caprine arthritis-encephalitis virus. *Am J Vet Res,* 44**,** 1670-5.

AHN, J., VU, T., NOVINCE, Z., GUERRERO-SANTORO, J., RAPIC-OTRIN, V. & GRONENBORN, A. M. 2010. HIV-1 Vpr Loads Uracil DNA Glycosylase-2 onto DCAF1, a Substrate Recognition Subunit of a Cullin 4A-RING E3 Ubiquitin Ligase for Proteasome-dependent Degradation*. *Journal of Biological Chemistry,* 285**,** 37333-37341.

AIYER, S., ROSSI, P., MALANI, N., SCHNEIDER, W. M., CHANDAR, A., BUSHMAN, F. D., MONTELIONE, G. T. & ROTH, M. J. 2015. Structural and sequencing analysis of local target DNA recognition by MLV integrase. *Nucleic acids research,* 43**,** 5647-5663.

AMARTELY, H., AVRAHAM, O., FRIEDLER, A., LIVNAH, O. & LEBENDIKER, M. 2018. Coupling Multi Angle Light Scattering to Ion Exchange chromatography (IEX-MALS) for protein characterization. *Scientific Reports,* 8**,** 6907.

ARHEL, N. J., SOUQUERE-BESSE, S., MUNIER, S., SOUQUE, P., GUADAGNINI, S., RUTHERFORD, S., PRÉVOST, M.-C., ALLEN, T. D. & CHARNEAU, P. 2007. HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. *The EMBO Journal,* 26**,** 3025-3037.

ARIYOSHI, M., VASSYLYEV, D. G., IWASAKI, H., NAKAMURA, H., SHINAGAWA, H. & MORIKAWA, K. 1994. Atomic structure of the RuvC resolvase: A holliday junction-specific endonuclease from E. coli. *Cell,* 78**,** 1063-1072.

ARNAUD, F., BLACK, S. G., MURPHY, L., GRIFFITHS, D. J., NEIL, S. J., SPENCER, T. E. & PALMARINI, M. 2010. Interplay between Ovine Bone Marrow Stromal Cell Antigen 2/Tetherin and Endogenous Retroviruses. *Journal of Virology,* 84**,** 4415.

BAE, S., PARK, J. & KIM, J.-S. 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics,* 30**,** 1473-1475.

BALLANDRAS-COLAS, A., BROWN, M., COOK, N. J., DEWDNEY, T. G., DEMELER, B., CHEREPANOV, P., LYUMKIS, D. & ENGELMAN, A. N. 2016. Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature,* 530**,** 358-61.

BALLANDRAS-COLAS, A., CHIVUKULA, V., SINGH, P., GRIFFITHS, D., ENGELMAN, A. & CHEREPANOV, P. Manuscript in preparation.

BALLANDRAS-COLAS, A., MASKELL, D. P., SERRAO, E., LOCKE, J., SWUEC, P., JONSSON, S. R., KOTECHA, A., COOK, N. J., PYE, V. E., TAYLOR, I. A., ANDRESDOTTIR, V., ENGELMAN, A. N., COSTA, A. & CHEREPANOV, P.

2017. A supramolecular assembly mediates lentiviral DNA integration. *Science,* 355**,** 93-95.

BALTIMORE, D. 1970. Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature,* 226**,** 1209-1211.

BALVAY, L., LOPEZ LASTRA, M., SARGUEIL, B., DARLIX, J. L. & OHLMANN, T. 2007. Translational control of retroviruses. *Nat Rev Microbiol,* 5**,** 128-40.

BARSKI, M. S., MINNELL, J. J., HODAKOVA, Z., PYE, V. E., NANS, A., CHEREPANOV, P. & MAERTENS, G. N. 2020. Cryo-EM structure of the deltaretroviral intasome in complex with the PP2A regulatory subunit B56γ. *Nature Communications,* 11**,** 5043.

BAUDE, A., AAES, T. L., ZHAI, B., AL-NAKOUZI, N., OO, H. Z., DAUGAARD, M., ROHDE, M. & JÄÄTTELÄ, M. 2016. Hepatoma-derived growth factor-related protein 2 promotes DNA repair by homologous recombination. *Nucleic acids research,* 44**,** 2214-2226.

BEDWELL, G. J. & ENGELMAN, A. N. 2021. Factors that mold the nuclear landscape of HIV-1 integration. *Nucleic Acids Research,* 49**,** 621-635.

BENLEULMI, M. S., MATYSIAK, J., HENRIQUEZ, D. R., VAILLANT, C., LESBATS, P., CALMELS, C., NAUGHTIN, M., LEON, O., SKALKA, A. M., RUFF, M., LAVIGNE, M., ANDREOLA, M.-L. & PARISSI, V. 2015. Intasome architecture and chromatin density modulate retroviral integration into nucleosome. *Retrovirology,* 12**,** 13-13.

BERRY, N., JAFFAR, S., VAN DER LOEFF, M. S., ARIYOSHI, K., HARDING, E., N'GOM, P. T., DIAS, F., WILKINS, A., RICARD, D., AABY, P., TEDDER, R. & WHITTLE, H. 2002. Low Level Viremia and High CD4% Predict Normal Survival in a Cohort of HIV Type-2-Infected Villagers. *AIDS Research and Human Retroviruses,* 18**,** 1167-1173.

BERTOLOTTI, L., MAZZEI, M., PUGGIONI, G., CARROZZA, M. L., DEI GIUDICI, S., MUZ, D., JUGANARU, M., PATTA, C., TOLARI, F. & ROSATI, S. 2011. Characterization of new small ruminant lentivirus subtype B3 suggests animal trade within the Mediterranean Basin. *Journal of General Virology,* 92**,** 1923-1929.

BHATT, V., SHI, K., SALAMANGO, D. J., MOELLER, N. H., PANDEY, K. K., BERA, S., BOHL, T. E., KURNIAWAN, F., ORELLANA, K., ZHANG, W., GRANDGENETT, D. P., HARRIS, R. S., SUNDBORGER-LUNNA, A. C. & AIHARA, H. 2020. Structural basis of host protein hijacking in human T-cell leukemia virus integration. *Nature Communications,* 11**,** 3121.

BLACKLAWS, B. A., BERRIATUA, E., TORSTEINSDOTTIR, S., WATT, N. J., DE ANDRES, D., KLEIN, D. & HARKISS, G. D. 2004. Transmission of small ruminant lentiviruses. *Veterinary Microbiology,* 101**,** 199-208.

BLUMENTHAL, R., DURELL, S. & VIARD, M. 2012. HIV Entry and Envelope Glycoprotein-mediated Fusion *. *Journal of Biological Chemistry,* 287**,** 40841-40849.

BORLONGAN, C. V. & EMERICH, D. F. 2010. Choroid Plexus and Immune Response of the Brain. *In:* BERCZI, I. (ed.) *New Insights to Neuroimmune Biology.* London: Elsevier.

BORSETTI, A., OHAGEN, A. & GÖTTLINGER, H. G. 1998. The C-terminal half of the human immunodeficiency virus type 1 Gag precursor is sufficient for efficient particle assembly. *Journal of virology,* 72**,** 9313-9317.

BOWERMAN, B., BROWN, P. O., BISHOP, J. M. & VARMUS, H. E. 1989. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev,* 3**,** 469-78.

BRIGGS, J. A. G., GRÜNEWALD, K., GLASS, B., FÖRSTER, F., KRÄUSSLICH, H.-G. & FULLER, S. D. 2006. The Mechanism of HIV-1 Core Assembly: Insights from Three-Dimensional Reconstructions of Authentic Virions. *Structure,* 14**,** 15-20.

BRIGGS, J. A. G., SIMON, M. N., GROSS, I., KRÄUSSLICH, H.-G., FULLER, S. D., VOGT, V. M. & JOHNSON, M. C. 2004. The stoichiometry of Gag protein in HIV-1. *Nature Structural & Molecular Biology,* 11**,** 672-675.

BRODER, S. & YARCHOAN, R. 1990. Dideoxycytidine: Current clinical experience and future prospects: A summary. *The American Journal of Medicine,* 88**,** S31-S33.

BROWN, K. M. & GILMARTIN, G. M. 2003. A Mechanism for the Regulation of Pre-mRNA 3′ Processing by Human Cleavage Factor Im. *Molecular Cell,* 12**,** 1467-1476.

BUKOVSKY, A. & GÖTTLINGER, H. 1996. Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease. *Journal of Virology,* 70**,** 6820.

BUKRINSKY, M. I., SHAROVA, N., MCDONALD, T. L., PUSHKARSKAYA, T., TARPLEY, W. G. & STEVENSON, M. 1993. Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 6125-6129.

BURKE, C. J., SANYAL, G., BRUNER, M. W., RYAN, J. A., LAFEMINA, R. L., ROBBINS, H. L., ZEFT, A. S., MIDDAUGH, C. R. & CORDINGLEY, M. G. 1992. Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *Journal of Biological Chemistry,* 267**,** 9639-9644.

BUSHMAN, F. D. & CRAIGIE, R. 1991. Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. *Proceedings of the National Academy of Sciences,* 88**,** 1339.

BUSSCHOTS, K., VERCAMMEN, J., EMILIANI, S., BENAROUS, R., ENGELBORGHS, Y., CHRIST, F. & DEBYSER, Z. 2005. The Interaction of LEDGF/p75 with Integrase Is Lentivirus-specific and Promotes DNA Binding *. *Journal of Biological Chemistry,* 280**,** 17841-17847.

BUSTIN, M. & REEVES, R. 1996. High-Mobility-Group Chromosomal Proteins: Architectural Components That Facilitate Chromatin Function. *In:* COHN, W. E. & MOLDAVE, K. (eds.) *Progress in Nucleic Acid Research and Molecular Biology.* Academic Press.

BUTLER, S. L., HANSEN, M. S. T. & BUSHMAN, F. D. 2001. A quantitative assay for HIV DNA integration in vivo. *Nature Medicine,* 7**,** 631-634.

BUTLER, S. L., JOHNSON, E. P. & BUSHMAN, F. D. 2002. Human Immunodeficiency Virus cDNA Metabolism: Notable Stability of Two-Long Terminal Repeat Circles. *Journal of Virology,* 76**,** 3739-3747.

CAI, M., ZHENG, R., CAFFREY, M., CRAIGIE, R., CLORE, G. M. & GRONENBORN, A. M. 1997. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Structural Biology,* 4**,** 567.

CARTEAU, S., HOFFMANN, C. & BUSHMAN, F. 1998. Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. *Journal of virology,* 72**,** 4005-4014.

CASE, K. 1986. Nomenclature: Human Immunodeficiency Virus. *Annals of Internal Medicine,* 105**,** 133-133.

CASWELL, S. J. 2019. *Biochemical analysis of the catalytic activity of HIV-1 restriction factor SAMHD1.*

CECCHERINI-SILBERSTEIN, F., MALET, I., FABENI, L., DIMONTE, S., SVICHER, V., D'ARRIGO, R., ARTESE, A., COSTA, G., BONO, S., ALCARO, S., D'ARMINIO MONFORTE, A., KATLAMA, C., CALVEZ, V., ANTINORI, A., MARCELIN, A.-G. & PERNO, C.-F. 2010. Specific HIV-1 integrase polymorphisms change their prevalence in untreated versus antiretroviral-treated HIV-1-infected patients, all naive to integrase inhibitors. *Journal of Antimicrobial Chemotherapy,* 65**,** 2305-2318.

CHAKRABARTI, L., GUYADER, M., ALIZON, M., DANIEL, M. D., DESROSIERS, R. C., TIOLLAIS, P. & SONIGO, P. 1987. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature,* 328**,** 543-547.

CHARMAN, H. P., BLADEN, S., GILDEN, R. V. & COGGINS, L. 1976. Equine infectious anemia virus: evidence favoring classification as a retravirus. *Journal of virology,* 19**,** 1073-1079.

CHARNEAU, P., ALIZON, M. & CLAVEL, F. 1992. A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *Journal of Virology,* 66**,** 2814.

CHELLAPPAN, S., KIRAN KUMAR REDDY, G. S., ALI, A., NALAM, M. N. L., ANJUM, S. G., CAO, H., KAIRYS, V., FERNANDES, M. X., ALTMAN, M. D., TIDOR, B., RANA, T. M., SCHIFFER, C. A. & GILSON, M. K. 2007. Design of Mutation-resistant HIV Protease Inhibitors with the Substrate Envelope Hypothesis. *Chemical Biology & Drug Design,* 69**,** 298-313.

CHEN, H. & ENGELMAN, A. 1998. The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 15270-15274.

CHEN, H., WEI, S. Q. & ENGELMAN, A. 1999. Multiple integrase functions are required to form the native structure of the human immunodeficiency virus type I intasome. *J Biol Chem,* 274**,** 17358-64.

CHEN, J. C., KRUCINSKI, J., MIERCKE, L. J., FINER-MOORE, J. S., TANG, A. H., LEAVITT, A. D. & STROUD, R. M. 2000. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc Natl Acad Sci U S A,* 97**,** 8233-8.

CHEREPANOV, P. 2007. LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res,* 35**,** 113-24.

CHEREPANOV, P., AMBROSIO, A. L. B., RAHMAN, S., ELLENBERGER, T. & ENGELMAN, A. 2005a. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 17308.

CHEREPANOV, P., DEVROE, E., SILVER, P. A. & ENGELMAN, A. 2004. Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J Biol Chem,* 279**,** 48883-92.

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**,** 372-81.

CHEREPANOV, P., SUN, Z. Y., RAHMAN, S., MAERTENS, G., WAGNER, G. & ENGELMAN, A. 2005b. Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat Struct Mol Biol,* 12**,** 526-32.

CHEREPANOV, P., SURRATT, D., TOELEN, J., PLUYMERS, W., GRIFFITH, J., DE CLERCQ, E. & DEBYSER, Z. 1999. Activity of recombinant HIV-1 integrase on mini-HIV DNA. *Nucleic Acids Research,* 27**,** 2202-2210.

CHEREPANOV, P. P. & WACKERNAGEL, W. 1995. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene,* 158**,** 9-14.

CHRIST, F., SHAW, S., DEMEULEMEESTER, J., DESIMMIE, B. A., MARCHAND, A., BUTLER, S., SMETS, W., CHALTIN, P., WESTBY, M., DEBYSER, Z. & PICKFORD, C. 2012. Small-Molecule Inhibitors of the LEDGF/p75 Binding Site of Integrase Block HIV Replication and Modulate Integrase Multimerization. *Antimicrobial Agents and Chemotherapy,* 56**,** 4365.

CHRIST, F., VOET, A., MARCHAND, A., NICOLET, S., DESIMMIE, B. A., MARCHAND, D., BARDIOT, D., VAN DER VEKEN, N. J., VAN REMOORTEL, B., STRELKOV, S. V., DE MAEYER, M., CHALTIN, P. & DEBYSER, Z. 2010. Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nature Chemical Biology,* 6**,** 442.

CIUFFI, A., LLANO, M., POESCHLA, E., HOFFMANN, C., LEIPZIG, J., SHINN, P., ECKER, J. R. & BUSHMAN, F. 2005. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med,* 11**,** 1287-1289.

CLARK, E. L., BUSH, S. J., MCCULLOCH, M. E. B., FARQUHAR, I. L., YOUNG, R., LEFEVRE, L., PRIDANS, C., TSANG, H. G., WU, C., AFRASIABI, C., WATSON, M., WHITELAW, C. B., FREEMAN, T. C., SUMMERS, K. M., ARCHIBALD, A. L. & HUME, D. A. 2017. A high resolution atlas of gene expression in the domestic sheep (Ovis aries). *PLOS Genetics,* 13**,** e1006997.

CLAVEL, F., GUETARD, D., BRUN-VEZINET, F., CHAMARET, S., REY, M. A., SANTOS-FERREIRA, M. O., LAURENT, A. G., DAUGUET, C., KATLAMA, C., ROUZIOUX, C. & AL, E. 1986. Isolation of a new human retrovirus from West African patients with AIDS. *Science,* 233**,** 343.

COFFIN, J., HAASE, A., LEVY, J. A., MONTAGNIER, L., OROSZLAN, S., TEICH, N., TEMIN, H., TOYOSHIMA, K., VARMUS, H., VOGT, P. & AL, E. 1986. Human immunodeficiency viruses. *Science,* 232**,** 697.

COFFIN, J. M., HUGHES, S. H. & VARMUS, H. E. (eds.) 1997. *Retroviruses,* Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

CONTICELLO, S. G., HARRIS, R. S. & NEUBERGER, M. S. 2003. The Vif Protein of HIV Triggers Degradation of the Human Antiretroviral DNA Deaminase APOBEC3G. *Current Biology,* 13**,** 2009-2013.

COOK, N. J., LI, W., BERTA, D., BADAOUI, M., BALLANDRAS-COLAS, A., NANS, A., KOTECHA, A., ROSTA, E., ENGELMAN, A. N. & CHEREPANOV, P. 2020. Structural basis of second-generation HIV integrase inhibitor action and viral resistance. *Science,* 367**,** 806.

COOPER, D. A., STEIGBIGEL, R. T., GATELL, J. M., ROCKSTROH, J. K., KATLAMA, C., YENI, P., LAZZARIN, A., CLOTET, B., KUMAR, P. N., ERON, J. E., SCHECHTER, M., MARKOWITZ, M., LOUTFY, M. R., LENNOX, J. L., ZHAO, J., CHEN, J., RYAN, D. M., RHODES, R. R., KILLAR, J. A., GILDE, L. R., STROHMAIER, K. M., MEIBOHM, A. R., MILLER, M. D., HAZUDA, D. J., NESSLY, M. L., DINUBILE, M. J., ISAACS, R. D., TEPPLER, H. & NGUYEN, B.-Y. 2008. Subgroup and Resistance Analyses of Raltegravir for Resistant HIV-1 Infection. *New England Journal of Medicine,* 359**,** 355-365.

COSNEFROY, O., MURRAY, P. J. & BISHOP, K. N. 2016. HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology,* 13**,** 58.

CRESPO, H., REINA, R., GLARIA, I., RAMÍREZ, H., DE ANDRÉS, X., JÁUREGUI, P., LUJÁN, L., MARTÍNEZ-POMARES, L., AMORENA, B. & DE ANDRÉS, D. F. 2011. Identification of the ovine mannose receptor and its possible role in Visna/Maedi virus infection. *Vet Res,* 42**,** 28-28.

CRISE, B., LI, Y., YUAN, C., MORCOCK, D. R., WHITBY, D., MUNROE, D. J., ARTHUR, L. O. & WU, X. 2005. Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. *Journal of virology,* 79**,** 12199-12204.

CULLEN, B. R. 1998. Retroviruses as Model Systems for the Study of Nuclear RNA Export Pathways. *Virology,* 249**,** 203-210.

CUTLIP, R. C. & LEHMKUHL, H. D. 1986. Eradication of ovine progressive pneumonia from sheep flocks. *J Am Vet Med Assoc,* 188**,** 1026-7.

DAMOND, F., DESCAMPS, D., FARFARA, I., TELLES, J. N., PUYEO, S., CAMPA, P., LEPRÊTRE, A., MATHERON, S., BRUN-VEZINET, F. & SIMON, F. 2001. Quantification of Proviral Load of Human Immunodeficiency Virus Type 2 Subtypes A and B Using Real-Time PCR. *Journal of Clinical Microbiology,* 39**,** 4264.

DAUGAARD, M., BAUDE, A., FUGGER, K., POVLSEN, L. K., BECK, H., SØRENSEN, C. S., PETERSEN, N. H. T., SORENSEN, P. H. B., LUKAS, C., BARTEK, J., LUKAS, J., ROHDE, M. & JÄÄTTELÄ, M. 2012. LEDGF (p75) promotes DNA-end resection and homologous recombination. *Nature Structural & Molecular Biology,* 19**,** 803-810.

DAVIES, J. F., HOSTOMSKA, Z., HOSTOMSKY, Z., JORDAN & MATTHEWS, D. A. 1991. Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science,* 252**,** 88.

DAWSON, M., VENABLES, C. & JENKINS, C. E. 1985. Experimental infection of a natural case of sheep pulmonary adenomatosis with maedi-visna virus. *Vet Rec,* 116**,** 588-9.

DAYTON, A. I., SODROSKI, J. G., ROSEN, C. A., GOH, W. C. & HASELTINE, W. A. 1986. The *trans*-activator gene of the human T cell lymphotropic virus type III is required for replication. *Cell,* 44**,** 941-947.

DE CLERCQ, E. 1995. Antiviral therapy for human immunodeficiency virus infections. *Clinical Microbiology Reviews,* 8**,** 200.

DE CLERCQ, E. 1998. The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection1Presented at the Eleventh International Conference on Antiviral Research, San Diego, CA, 5–10 April 1998.1. *Antiviral Research,* 38**,** 153-179.

DE RIJCK, J., DE KOGEL, C., DEMEULEMEESTER, J., VETS, S., EL ASHKAR, S., MALANI, N., BUSHMAN, F. D., LANDUYT, B., HUSSON, S. J., BUSSCHOTS, K., GIJSBERS, R. & DEBYSER, Z. 2013. The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep,* 5**,** 886-894.

DELELIS, O., CARAYON, K., GUIOT, E., LEH, H., TAUC, P., BROCHON, J.-C., MOUSCADET, J.-F. & DEPREZ, E. 2008. Insight into the Integrase-DNA Recognition Mechanism: A Specific DNA-binding mode revealed by an enzymatically labeled Integrase*. *Journal of Biological Chemistry,* 283**,** 27838-27849.

DEMEULEMEESTER, J., VETS, S., SCHRIJVERS, R., MADLALA, P., DE MAEYER, M., DE RIJCK, J., NDUNG'U, T., DEBYSER, Z. & GIJSBERS, R. 2014. HIV-1 Integrase Variants Retarget Viral Integration and Are Associated with Disease Progression in a Chronic Infection Cohort. *Cell Host & Microbe,* 16**,** 651-662.

DENG, H., LIU, R., ELLMEIER, W., CHOE, S., UNUTMAZ, D., BURKHART, M., MARZIO, P. D., MARMON, S., SUTTON, R. E., HILL, C. M., DAVIS, C. B., PEIPER, S. C., SCHALL, T. J., LITTMAN, D. R. & LANDAU, N. R. 1996. Identification of a major co-receptor for primary isolates of HIV-1. *Nature,* 381**,** 661-666.

DERSE, D., CRISE, B., LI, Y., PRINCLER, G., LUM, N., STEWART, C., MCGRATH, C. F., HUGHES, S. H., MUNROE, D. J. & WU, X. 2007. Human T-cell leukemia virus type 1 integration target sites in the human genome: comparison with those of other retroviruses. *Journal of virology,* 81**,** 6731-6741.

DEVITA, V. T., BRODER, S., FAUCI, A. S., KOVACS, J. A. & CHABNER, B. A. 1987. Developmental Therapeutics and the Acquired Immunodeficiency Syndrome. *Annals of Internal Medicine,* 106**,** 568-581.

DICK, R. A., ZADROZNY, K. K., XU, C., SCHUR, F. K. M., LYDDON, T. D., RICANA, C. L., WAGNER, J. M., PERILLA, J. R., GANSER-PORNILLOS, B. K., JOHNSON, M. C., PORNILLOS, O. & VOGT, V. M. 2018. Inositol phosphates are assembly co-factors for HIV-1. *Nature,* 560**,** 509-512.

DUKES, T. W., GREIG, A. S. & CORNER, A. H. 1979. Maedi-visna in Canadian sheep. *Can J Comp Med,* 43**,** 313-20.

DYDA, F., HICKMAN, A. B., JENKINS, T. M., ENGELMAN, A., CRAIGIE, R. & DAVIES, D. R. 1994. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science,* 266**,** 1981.

EIJKELENBOOM, A. P. A. M., PURAS LUTZKE, R. A., BOELENS, R., PLASTERK, R. H. A., KAPTEIN, R. & HÅRD, K. 1995. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nature Structural Biology,* 2**,** 807.

ELLIOTT, J. L., ESCHBACH, J. E., KONERU, P. C., LI, W., PURAY-CHAVEZ, M., TOWNSEND, D., LAWSON, D. Q., ENGELMAN, A. N., KVARATSKHELIA, M. & KUTLUAY, S. B. 2020. Integrase-RNA interactions underscore the critical role of integrase in HIV-1 virion morphogenesis. *eLife,* 9**,** e54311.

EMERMAN, M., VAZEUX, R. & PEDEN, K. 1989. The *rev* gene product of the human immunodeficiency virus affects envelope-specific RNA localization. *Cell,* 57**,** 1155-1165.

EMILIANI, S., MOUSNIER, A., BUSSCHOTS, K., MAROUN, M., VAN MAELE, B., TEMPÉ, D., VANDEKERCKHOVE, L., MOISANT, F., BEN-SLAMA, L., WITVROUW, M., CHRIST, F., RAIN, J.-C., DARGEMONT, C., DEBYSER, Z. & BENAROUS, R. 2005. Integrase Mutants Defective for Interaction with LEDGF/p75 Are Impairedin Chromosome Tethering and HIV-1Replication*. *Journal of Biological Chemistry,* 280**,** 25517-25523.

ENGELMAN, A. 1999. In Vivo Analysis of Retroviral Integrase Structure and Function. *In:* RLARAMOROSCH, K., MURPHY, F. A. & SHAWN, A. J. (eds.) *Advances in Virus Research.* Academic Press.

ENGELMAN, A. & CRAIGIE, R. 1992. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *Journal of Virology,* 66**,** 6361-6369.

ENGELMAN, A. & CRAIGIE, R. 1995. Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity. *Journal of virology,* 69**,** 5908-5911.

ENGELMAN, A., ENGLUND, G., ORENSTEIN, J. M., MARTIN, M. A. & CRAIGIE, R. 1995. Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *Journal of virology,* 69**,** 2729-2736.

ENGELMAN, A., MIZUUCHI, K. & CRAIGIE, R. 1991. HIV-1 DNA integration: Mechanism of viral DNA cleavage and DNA strand transfer. *Cell,* 67**,** 1211-1221.

ENGELMAN, A. N. 2019. Multifaceted HIV integrase functionalities and therapeutic strategies for their inhibition. *The Journal of biological chemistry,* 294**,** 15137-15157.

ENGELMAN, A. N. & CHEREPANOV, P. 2017. Retroviral intasomes arising. *Curr Opin Struct Biol,* 47**,** 23-29.

ENGELMAN, A. N. & CHEREPANOV, P. 2021. Close-up: HIV/SIV intasome structures shed new light on integrase inhibitor binding and viral escape mechanisms. *The FEBS Journal,* 288**,** 427-433.

ESPESETH, A. S., FELOCK, P., WOLFE, A., WITMER, M., GROBLER, J., ANTHONY, N., EGBERTSON, M., MELAMED, J. Y., YOUNG, S., HAMILL, T., COLE, J. L. & HAZUDA, D. J. 2000. HIV-1 integrase inhibitors that compete with the target DNA

substrate define a unique strand transfer conformation for integrase. *Proceedings of the National Academy of Sciences,* 97**,** 11244-11249.

FADEL, H. J., MORRISON, J. H., SAENZ, D. T., FUCHS, J. R., KVARATSKHELIA, M., EKKER, S. C. & POESCHLA, E. M. 2014. TALEN Knockout of the *PSIP1* Gene in Human Cells: Analyses of HIV-1 Replication and Allosteric Integrase Inhibitor Mechanism. *Journal of Virology,* 88**,** 9704.

FARNET, C. M. & BUSHMAN, F. D. 1997. HIV-1 cDNA Integration: Requirement of HMG I(Y) Protein for Function of Preintegration Complexes In Vitro. *Cell,* 88**,** 483-492.

FASHENA, S. J., REEVES, R. & RUDDLE, N. H. 1992. A poly(dA-dT) upstream activating sequence binds high-mobility group I protein and contributes to lymphotoxin (tumor necrosis factor-beta) gene regulation. *Molecular and Cellular Biology,* 12**,** 894.

FAURE, A., CALMELS, C., DESJOBERT, C., CASTROVIEJO, M., CAUMONT-SARCOS, A., TARRAGO-LITVAK, L., LITVAK, S. & PARISSI, V. 2005. HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res,* 33**,** 977-86.

FELBER, B. K., HADZOPOULOU-CLADARAS, M., CLADARAS, C., COPELAND, T. & PAVLAKIS, G. N. 1989. rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proceedings of the National Academy of Sciences,* 86**,** 1495.

FENG, L., DHARMARAJAN, V., SERRAO, E., HOYTE, A., LARUE, R. C., SLAUGHTER, A., SHARMA, A., PLUMB, M. R., KESSL, J. J., FUCHS, J. R., BUSHMAN, F. D., ENGELMAN, A. N., GRIFFIN, P. R. & KVARATSKHELIA, M. 2016. The Competitive Interplay between Allosteric HIV-1 Integrase Inhibitor BI/D and LEDGF/p75 during the Early Stage of HIV-1 Replication Adversely Affects Inhibitor Potency. *ACS Chemical Biology,* 11**,** 1313-1321.

FENG, Y., BRODER, C. C., KENNEDY, P. E. & BERGER, E. A. 1996. HIV-1 Entry Cofactor: Functional cDNA Cloning of a Seven-Transmembrane, G Protein-Coupled Receptor. *Science,* 272**,** 872-877.

FISHER, A. G., FEINBERG, M. B., JOSEPHS, S. F., HARPER, M. E., MARSELLE, L. M., REYES, G., GONDA, M. A., ALDOVINI, A., DEBOUK, C., GALLO, R. C. & WONG-STAAL, F. 1986. The trans-activator gene of HTLV-III is essential for virus replication. *Nature,* 320**,** 367-371.

FONTANA, J., JURADO, K. A., CHENG, N., LY, N. L., FUCHS, J. R., GORELICK, R. J., ENGELMAN, A. N. & STEVEN, A. C. 2015. Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-Inhibited Virions: a Role for Integrase in Maturation. *Journal of Virology,* 89**,** 9765.

FRANCIS, A. C., MARIN, M., PRELLBERG, M. J., PALERMINO-ROWLAND, K. & MELIKYAN, G. B. 2020. HIV-1 Uncoating and Nuclear Import Precede the Completion of Reverse Transcription in Cell Lines and in Primary Macrophages. *Viruses,* 12**,** 1234.

FRANCIS, A. C. & MELIKYAN, G. B. 2018. Single HIV-1 Imaging Reveals Progression of Infection through CA-Dependent Steps of Docking at the Nuclear Pore, Uncoating, and Nuclear Transport. *Cell Host & Microbe,* 23**,** 536-548.e6.

FURMAN, P. A., FYFE, J. A., ST CLAIR, M. H., WEINHOLD, K., RIDEOUT, J. L., FREEMAN, G. A., LEHRMAN, S. N., BOLOGNESI, D. P., BRODER, S.,

MITSUYA, H. & ET AL. 1986. Phosphorylation of 3′-azido-3′-deoxythymidine and selective interaction of the 5′-triphosphate with human immunodeficiency virus reverse transcriptase. *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 8333-8337.

GALLAY, P., SWINGLER, S., SONG, J., BUSHMAN, F. & TRONO, D. 1995. HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell,* 83**,** 569-576.

GALLO, R. C., SALAHUDDIN, S. Z., POPOVIC, M., SHEARER, G. M., KAPLAN, M., HAYNES, B. F., PALKER, T. J., REDFIELD, R., OLESKE, J., SAFAI, B. & ET, A. 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science,* 224**,** 500.

GAO, F., BAILES, E., ROBERTSON, D. L., CHEN, Y., RODENBURG, C. M., MICHAEL, S. F., CUMMINS, L. B., ARTHUR, L. O., PEETERS, M., SHAW, G. M., SHARP, P. M. & HAHN, B. H. 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature,* 397**,** 436-441.

GAO, F., YUE, L., WHITE, A. T., PAPPAS, P. G., BARCHUE, J., HANSON, A. P., GREENE, B. M., SHARP, P. M., SHAW, G. M. & HAHN, B. H. 1992. Human infection by genetically diverse SIVSM-related HIV-2 in West Africa. *Nature,* 358**,** 495-499.

GARCIA, J. V. & MILLER, A. D. 1991. Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature,* 350**,** 508-511.

GARRUS, J. E., VON SCHWEDLER, U. K., PORNILLOS, O. W., MORHAM, S. G., ZAVITZ, K. H., WANG, H. E., WETTSTEIN, D. A., STRAY, K. M., CÔTÉ, M., RICH, R. L., MYSZKA, D. G. & SUNDQUIST, W. I. 2001. Tsg101 and the Vacuolar Protein Sorting Pathway Are Essential for HIV-1 Budding. *Cell,* 107**,** 55-65.

GE, H., SI, Y. & ROEDER, R. G. 1998. Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *The EMBO Journal,* 17**,** 6723-6729.

GENDELMAN, H. E., NARAYAN, O., KENNEDY-STOSKOPF, S., KENNEDY, P. G., GHOTBI, Z., CLEMENTS, J. E., STANLEY, J. & PEZESHKPOUR, G. 1986. Tropism of sheep lentiviruses for monocytes: susceptibility to infection and virus gene expression increase during maturation of monocytes to macrophages. *J Virol,* 58**,** 67-74.

GESSAIN, A., VERNANT, J. C., MAURS, L., BARIN, F., GOUT, O., CALENDER, A. & DE THÉ, G. 1985. Antibodies to human T-Lymphotropic virus type-I in patients with tropical spastic paraparesis. *The Lancet,* 326**,** 407-410.

GHEYSEN, D., JACOBS, E., DE FORESTA, F., THIRIART, C., FRANCOTTE, M., THINES, D. & DE WILDE, M. 1989. Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. *Cell,* 59**,** 103-112.

GIAMMARIOLI, M., BAZZUCCHI, M., PUGGIONI, G., BRAJON, G., DEI GIUDICI, S., TACCORI, F., FELIZIANI, F. & DE MIA, G. M. 2011. Phylogenetic analysis of small ruminant lentivirus (SRLV) in Italian flocks reveals the existence of novel genetic subtypes. *Virus Genes,* 43**,** 380-384.

GIBBONS, J. M., MARNO, K. M., PIKE, R., LEE, W.-Y. J., JONES, C. E., OGUNKOLADE, B. W., PARDIEU, C., BRYAN, A., FU, R. M., WARNES, G., ROWLEY, P. A., SLOAN, R. D. & MCKNIGHT, Á. 2020. HIV-1 Accessory Protein Vpr Interacts with REAF/RPRD2 To Mitigate Its Antiviral Activity. *Journal of Virology,* 94**,** e01591-19.

GJERSET, B., STORSET, A. K. & RIMSTAD, E. 2006. Genetic diversity of small-ruminant lentiviruses: characterization of Norwegian isolates of Caprine arthritis encephalitis virus. *Journal of General Virology,* 87**,** 573-580.

GOEPFERT, P. A., WANG, G. & MULLIGAN, M. J. 1995. Identification of an ER retrieval signal in a retroviral glycoprotein. *Cell,* 82**,** 543-544.

GREGO, E., BERTOLOTTI, L., QUASSO, A., PROFITI, M., LACERENZA, D., MUZ, D. & ROSATI, S. 2007. Genetic characterization of small ruminant lentivirus in Italian mixed flocks: Evidence for a novel genotype circulating in a local goat population. *Journal of General Virology,* 88**,** 3423-3427.

GROBLER, J. A., STILLMOCK, K., HU, B., WITMER, M., FELOCK, P., ESPESETH, A. S., WOLFE, A., EGBERTSON, M., BOURGEOIS, M., MELAMED, J., WAI, J. S., YOUNG, S., VACCA, J. & HAZUDA, D. J. 2002. Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes. *Proceedings of the National Academy of Sciences,* 99**,** 6661.

GRUBER, A. R., MARTIN, G., KELLER, W. & ZAVOLAN, M. 2012. Cleavage factor Im is a key regulator of 3′ UTR length. *RNA Biology,* 9**,** 1405-1412.

GUAN, R., AIYER, S., COTE, M. L., XIAO, R., JIANG, M., ACTON, T. B., ROTH, M. J. & MONTELIONE, G. T. 2017. X-ray crystal structure of the N-terminal region of Moloney murine leukemia virus integrase and its implications for viral DNA recognition. *Proteins,* 85**,** 647-656.

GUELEN, L., PAGIE, L., BRASSET, E., MEULEMAN, W., FAZA, M. B., TALHOUT, W., EUSSEN, B. H., DE KLEIN, A., WESSELS, L., DE LAAT, W. & VAN STEENSEL, B. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature,* 453**,** 948-951.

GUPTA, K., ALLEN, A., GIRALDO, C., EILERS, G., SHARP, R., HWANG, Y., MURALI, H., CRUZ, K., JANMEY, P., BUSHMAN, F. & VAN DUYNE, G. D. 2021. Allosteric HIV Integrase Inhibitors Promote Formation of Inactive Branched Polymers via Homomeric Carboxy-Terminal Domain Interactions. *Structure,* 29**,** 213-225.e5.

GUPTA, K., CURTIS, J. E., KRUEGER, S., HWANG, Y., CHEREPANOV, P., BUSHMAN, F. D. & VAN DUYNE, G. D. 2012. Solution conformations of prototype foamy virus integrase and its stable synaptic complex with U5 viral DNA. *Structure,* 20**,** 1918-28.

GUPTA, K., TURKKI, V., SHERRILL-MIX, S., HWANG, Y., EILERS, G., TAYLOR, L., MCDANAL, C., WANG, P., TEMELKOFF, D., NOLTE, R. T., VELTHUISEN, E., JEFFREY, J., VAN DUYNE, G. D. & BUSHMAN, F. D. 2016. Structural Basis for Inhibitor-Induced Aggregation of HIV Integrase. *PLOS Biology,* 14**,** e1002584.

GUPTA, S. S., MAETZIG, T., MAERTENS, G. N., SHARIF, A., ROTHE, M., WEIDNER-GLUNDE, M., GALLA, M., SCHAMBACH, A., CHEREPANOV, P. & SCHULZ, T. F. 2013. Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J Virol,* 87**,** 12721-36.

GUYADER, M., EMERMAN, M., SONIGO, P., CLAVEL, F., MONTAGNIER, L. & ALIZON, M. 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature,* 326**,** 662-669.

HAASE, A. T. 1975. The Slow Infection Caused by Visna Virus. *In:* ARBER, W., HENLE, W., HOFSCHNEIDER, P. H., HUMPHREY, J. H., JERNE, N. K., KOLDOVSKÝ, P., KOPROWSKI, H., MAALØE, O., ROTT, R., SCHWEIGER, H. G., SELA, M., SYRUČEK, L. & VOGT, P. K. (eds.) *Current Topics in Microbiology and Immunology: Volume 72.* Berlin, Heidelberg: Springer Berlin Heidelberg.

HAASE, A. T. 1986. Pathogenesis of lentivirus infections. *Nature,* 322**,** 130-6.

HAASE, A. T., STOWRING, L., HARRIS, J. D., TRAYNOR, B., VENTURA, P., PELUSO, R. & BRAHIC, M. 1982. Visna DNA synthesis and the tempo of infection in vitro. *Virology,* 119**,** 399-410.

HACKER, C. V., VINK, C. A., WARDELL, T. W., LEE, S., TREASURE, P., KINGSMAN, S. M., MITROPHANOUS, K. A. & MISKIN, J. E. 2006. The integration profile of EIAV-based vectors. *Molecular Therapy,* 14**,** 536-545.

HARE, S., DI NUNZIO, F., LABEJA, A., WANG, J., ENGELMAN, A. & CHEREPANOV, P. 2009a. Structural basis for functional tetramerization of lentiviral integrase. *PLoS Pathog,* 5**,** e1000515.

HARE, S., GUPTA, S. S., VALKOV, E., ENGELMAN, A. & CHEREPANOV, P. 2010. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature,* 464**,** 232-6.

HARE, S., MAERTENS, G. N. & CHEREPANOV, P. 2012. 3′-processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO J,* 31**,** 3020-8.

HARE, S., SHUN, M. C., GUPTA, S. S., VALKOV, E., ENGELMAN, A. & CHEREPANOV, P. 2009b. A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. *PLoS Pathog,* 5**,** e1000259.

HARE, S., SMITH, S. J., MÉTIFIOT, M., JAXA-CHAMIEC, A., POMMIER, Y., HUGHES, S. H. & CHEREPANOV, P. 2011. Structural and Functional Analyses of the Second-Generation Integrase Strand Transfer Inhibitor Dolutegravir (S/GSK1349572). *Molecular Pharmacology,* 80**,** 565.

HARRIS, J. D., BLUM, H., SCOTT, J., TRAYNOR, B., VENTURA, P. & HAASE, A. 1984. Slow virus visna: reproduction in vitro of virus from extrachromosomal DNA. *Proceedings of the National Academy of Sciences,* 81**,** 7212.

HAZUDA, D., BLAU, C. U., FELOCK, P., HASTINGS, J., PRAMANIK, B., WOLFE, A., BUSHMAN, F., FARNET, C., GOETZ, M., WILLIAMS, M., SILVERMAN, K., LINGHAM, R. & SINGH, S. 1999. Isolation and Characterization of Novel Human Immunodeficiency Virus Integrase Inhibitors from Fungal Metabolites. *Antiviral Chemistry and Chemotherapy,* 10**,** 63-70.

HAZUDA, D. J., FELOCK, P. J., HASTINGS, J. C., PRAMANIK, B. & WOLFE, A. L. 1997. Differential divalent cation requirements uncouple the assembly and catalytic reactions of human immunodeficiency virus type 1 integrase. *Journal of Virology,* 71**,** 7005.

HEHL, E. A., JOSHI, P., KALPANA, G. V. & PRASAD, V. R. 2004. Interaction between Human Immunodeficiency Virus Type 1 Reverse Transcriptase and Integrase Proteins. *Journal of Virology,* 78**,** 5056-5067.

HEINZINGER, N. K., BUKRINSKY, M. I., HAGGERTY, S. A., RAGLAND, A. M., KEWALRAMANI, V., LEE, M. A., GENDELMAN, H. E., RATNER, L., STEVENSON, M. & EMERMAN, M. 1994. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 7311-7315.

HIDER, R. C. & HODGES, S. J. 1984. Protein secondary structure: Analysis and prediction. *Biochemical Education,* 12**,** 9-18.

HIGHTOWER, K. E., WANG, R., DEANDA, F., JOHNS, B. A., WEAVER, K., SHEN, Y., TOMBERLIN, G. H., CARTER, H. L., BRODERICK, T., SIGETHY, S., SEKI, T., KOBAYASHI, M. & UNDERWOOD, M. R. 2011. Dolutegravir (S/GSK1349572) Exhibits Significantly Slower Dissociation than Raltegravir and Elvitegravir from Wild-Type and Integrase Inhibitor-Resistant HIV-1 Integrase-DNA Complexes. *Antimicrobial Agents and Chemotherapy,* 55**,** 4552.

HINUMA, Y., NAGATA, K., HANAOKA, M., NAKAI, M., MATSUMOTO, T., KINOSHITA, K. I., SHIRAKAWA, S. & MIYOSHI, I. 1981. Adult T-cell leukemia: antigen in an ATL cell line and detection of antibodies to the antigen in human sera. *Proceedings of the National Academy of Sciences,* 78**,** 6476.

HOLMAN, A. G. & COFFIN, J. M. 2005. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 6103-6107.

HOUWERS, D. J., KÖNIG, C. D. W., DE BOER, G. F. & SCHAAKE, J. 1983. Maedi-visna control in sheep I. Artificial rearing of colostrum-deprived lambs. *Veterinary Microbiology,* 8**,** 179-185.

HOVDEN, A.-O. & SOMMERFELT, M. A. 2002. The influence of CD4 and CXCR4 on maedi-visna virus-induced syncytium formation. *APMIS,* 110**,** 697-708.

HUANG, P., FARQUHAR, D. & PLUNKETT, W. 1990. Selective action of 3′-azido-3′-deoxythymidine 5′-triphosphate on viral reverse transcriptases and human DNA polymerases. *Journal of Biological Chemistry,* 265**,** 11914-11918.

HUGHES, S. H. 2015. Reverse Transcription of Retroviruses and LTR Retrotransposons. *Microbiology Spectrum,* 3.

HULME, A. E., PEREZ, O. & HOPE, T. J. 2011. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proceedings of the National Academy of Sciences,* 108**,** 9975.

IGAKURA, T., STINCHCOMBE, J. C., GOON, P. K. C., TAYLOR, G. P., WEBER, J. N., GRIFFITHS, G. M., TANAKA, Y., OSAME, M. & BANGHAM, C. R. M. 2003. Spread of HTLV-I Between Lymphocytes by Virus-Induced Polarization of the Cytoskeleton. *Science,* 299**,** 1713.

ISEL, C., EHRESMANN, C. & MARQUET, R. 2010. Initiation of HIV Reverse Transcription. *Viruses,* 2**,** 213-243.

JANG, S., COOK, N. J., PYE, V. E., BEDWELL, G. J., DUDEK, A. M., SINGH, P. K., CHEREPANOV, P. & ENGELMAN, A. N. 2019. Differential role for phosphorylation in alternative polyadenylation function versus nuclear import of SR-like protein CPSF6. *Nucleic Acids Res,* 47**,** 4663-4683.

JURADO, K. A., WANG, H., SLAUGHTER, A., FENG, L., KESSL, J. J., KOH, Y., WANG, W., BALLANDRAS-COLAS, A., PATEL, P. A., FUCHS, J. R., KVARATSKHELIA, M. & ENGELMAN, A. 2013. Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation. *Proceedings of the National Academy of Sciences,* 110**,** 8690.

KALOGIANNI, A. I., BOSSIS, I., EKATERINIADOU, L. V. & GELASAKIS, A. I. 2020. Etiology, Epizootiology and Control of Maedi-Visna in Dairy Sheep: A Review. *Animals (Basel),* 10**,** 616.

KALPANA, G. V., REICIN, A., CHENG, G. S. W., SORIN, M., PAIK, S. & GOFF, S. P. 1999. Isolation and Characterization of an Oligomerization-Negative Mutant of HIV-1 Integrase. *Virology,* 259**,** 274-285.

KALYANARAMAN, V. S., SARNGADHARAN, M. G., POIESZ, B., RUSCETTI, F. W. & GALLO, R. C. 1981. Immunological properties of a type C retrovirus isolated from cultured human T-lymphoma cells and comparison to other mammalian retroviruses. *Journal of Virology,* 38**,** 906.

KALYANARAMAN, V. S., SARNGADHARAN, M. G., ROBERT-GUROFF, M., MIYOSHI, I., GOLDE, D. & GALLO, R. C. 1982. A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. *Science,* 218**,** 571.

KANG, Y., MORESSI, C. J., SCHEETZ, T. E., XIE, L., TRAN, D. T., CASAVANT, T. L., AK, P., BENHAM, C. J., DAVIDSON, B. L. & MCCRAY, P. B., JR. 2006. Integration site choice of a feline immunodeficiency virus vector. *Journal of virology,* 80**,** 8820-8823.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M., HAUSSLER & DAVID 2002. The Human Genome Browser at UCSC. *Genome Res,* 12**,** 996-1006.

KESSL, J. J., KUTLUAY, S. B., TOWNSEND, D., REBENSBURG, S., SLAUGHTER, A., LARUE, R. C., SHKRIABAI, N., BAKOUCHE, N., FUCHS, J. R., BIENIASZ, P. D. & KVARATSKHELIA, M. 2016. HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell,* 166**,** 1257-1268.e12.

KHAN, E., MACK, J. P., KATZ, R. A., KULKOSKY, J. & SKALKA, A. M. 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Research,* 19**,** 851-860.

KHAN, H., SUMNER, R. P., RASAIYAAH, J., TAN, C. P., RODRIGUEZ-PLATA, M. T., VAN TULLEKEN, C., FINK, D., ZULIANI-ALVAREZ, L., THORNE, L., STIRLING, D., MILNE, R. S. B. & TOWERS, G. J. 2020. HIV-1 Vpr antagonizes innate immune activation by targeting karyopherin-mediated NF-κB/IRF3 nuclear transport. *eLife,* 9**,** e60821.

KIM, S., KIM, D., CHO, S. W., KIM, J. & KIM, J.-S. 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res,* 24**,** 1012-1019.

KING, A. M., ADAMS, M. J., CARSTENS, E. B. & LEFKOWITZ, E. B. 2011. *Virus Taxonomy : Ninth Report of the International Committee on Taxonomy of Virst. Louis, Mo, Usaes,* St. Louis, UNITED STATES, Elsevier.

KOH, Y., WU, X., FERRIS, A. L., MATREYEK, K. A., SMITH, S. J., LEE, K., KEWALRAMANI, V. N., HUGHES, S. H. & ENGELMAN, A. 2013. Differential effects of human immunodeficiency virus type 1 capsid and cellular factors nucleoporin 153 and LEDGF/p75 on the efficiency and specificity of viral DNA integration. *Journal of virology,* 87**,** 648-658.

KOHLSTAEDT, L. A., WANG, J., FRIEDMAN, J. M., RICE, P. A. & STEITZ, T. A. 1992. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science,* 256**,** 1783.

KRÄUSSLICH, H. G., FÄCKE, M., HEUSER, A. M., KONVALINKA, J. & ZENTGRAF, H. 1995. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *Journal of virology,* 69**,** 3407-3419.

KRISHNAN, L., LI, X., NARAHARISETTY, H. L., HARE, S., CHEREPANOV, P. & ENGELMAN, A. 2010. Structure-based modeling of the functional HIV-1 intasome and its inhibition. *Proc Natl Acad Sci U S A,* 107**,** 15910-5.

KRISTBJÖRNSDÓTTIR, H. B. S., ANDRÉSDÓTTIR, V., SVANSSON, V., TORSTEINSDÓTTIR, S., MATTHÍASDÓTTIR, S. D. & ANDRÉSSON, Ó. S. 2004. The vif gene of maedi-visna virus is essential for infectivity in vivo and in vitro. *Virology,* 318**,** 350-359.

KULKOSKY, J., JONES, K. S., KATZ, R. A., MACK, J. P. & SKALKA, A. M. 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Molecular and Cellular Biology,* 12**,** 2331-2338.

KUZEMBAYEVA, M., DILLEY, K., SARDO, L. & HU, W.-S. 2014. Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology,* 454-455**,** 362-370.

KWONG, P. D., WYATT, R., ROBINSON, J., SWEET, R. W., SODROSKI, J. & HENDRICKSON, W. A. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature,* 393**,** 648-659.

LAMA, J., MANGASARIAN, A. & TRONO, D. 1999. Cell-surface expression of CD4 reduces HIV-1 infectivity by blocking Env incorporation in a Nef- and Vpu-inhibitable manner. *Current Biology,* 9**,** 622-631.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, N., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER,

S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J.-F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LEAVITT, A. D., ROBLES, G., ALESANDRO, N. & VARMUS, H. E. 1996. Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection. *Journal of Virology,* 70**,** 721-728.

LEE, J. S., KILDEGAARD, H. F., LEWIS, N. E. & LEE, G. M. 2019. Mitigating Clonal Variation in Recombinant Mammalian Cell Lines. *Trends in Biotechnology,* 37**,** 931-942.

LEE, M. S. & CRAIGIE, R. 1994. Protection of retroviral DNA from autointegration: involvement of a cellular factor. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 9823-9827.

LEE, M. S. & CRAIGIE, R. 1998. A previously unidentified host protein protects retroviral DNA from autointegration. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 1528-1533.

LERONDELLE, C. & OUZROUT, R. 1990. Expression of maedi-visna virus in mammary secretions of a seropositive ewe. *Dev Biol Stand,* 72**,** 223-7.

LEROY, G., OKSUZ, O., DESCOSTES, N., AOI, Y., GANAI, R. A., KARA, H. O., YU, J.-R., LEE, C.-H., STAFFORD, J., SHILATIFARD, A. & REINBERG, D. 2019. LEDGF and HDGF2 relieve the nucleosome-induced barrier to transcription in differentiated cells. *Science Advances,* 5**,** eaay3068.

LESBATS, P., ENGELMAN, A. N. & CHEREPANOV, P. 2016. Retroviral DNA Integration. *Chem Rev,* 116**,** 12730-12757.

LEWIS, P. F. & EMERMAN, M. 1994. Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. *J Virol,* 68**,** 510-6.

LI, C., BURDICK, R. C., NAGASHIMA, K., HU, W.-S. & PATHAK, V. K. 2021. HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proceedings of the National Academy of Sciences,* 118**,** e2019467118.

LI, M., JURADO, K. A., LIN, S., ENGELMAN, A. & CRAIGIE, R. 2014. Engineered Hyperactive Integrase for Concerted HIV-1 DNA Integration. *PLOS ONE,* 9**,** e105078.

LI, W., LEE, M.-H., HENDERSON, L., TYAGI, R., BACHANI, M., STEINER, J., CAMPANAC, E., HOFFMAN, D. A., VON GELDERN, G., JOHNSON, K., MARIC, D., MORRIS, H. D., LENTZ, M., PAK, K., MAMMEN, A., OSTROW, L., ROTHSTEIN, J. & NATH, A. 2015. Human endogenous retrovirus-K contributes to motor neuron disease. *Science Translational Medicine,* 7**,** 307ra153.

LI, W., SINGH, P. K., SOWD, G. A., BEDWELL, G. J., JANG, S., ACHUTHAN, V., OLERU, A. V., WONG, D., FADEL, H. J., LEE, K., KEWALRAMANI, V. N., POESCHLA, E. M., HERSCHHORN, A. & ENGELMAN, A. N. 2020. CPSF6-Dependent Targeting of Speckle-Associated Domains Distinguishes Primate from Nonprimate Lentiviral Integration. *mBio,* 11**,** e02254-20.

LIGHT, M. R., SCHIPPER, I. A., MOLITOR, T. W., TILTON, J. E. & SLANGER, W. D. 1979. Progressive Pneumonia in Sheep: Incidence of Natural Infection and Establishment of Clean Flocks. *Journal of Animal Science,* 49**,** 1157-1160.

LIN, C.-W. & ENGELMAN, A. 2003. The Barrier-to-Autointegration Factor Is a Component of Functional Human Immunodeficiency Virus Type 1 Preintegration Complexes. *Journal of Virology,* 77**,** 5030.

LIN, J. J. 1992. Endonuclease A degrades chromosomal and plasmid DNA of Escherichia coli present in most preparations of single stranded DNA from phagemids. *Proc Natl Sci Counc Repub China B,* 16**,** 1-5.

LINGAPPA, J. R., REED, J. C., TANAKA, M., CHUTIRAKA, K. & ROBINSON, B. A. 2014. How HIV-1 Gag assembles in cells: Putting together pieces of the puzzle. *Virus Research,* 193**,** 89-107.

LIST, J. & HAASE, A. T. 1997. Integration of Visna Virus DNA Occurs and May Be Necessary for Productive Infection. *Virology,* 237**,** 189-197.

LITVAK, S., SARIH-COTTIN, L., FOURNIER, M., ANDREOLA, M. & TARRAGO-LITVAK, L. 1994. Priming of HIV replication by tRNALys3: role of reverse transcriptase. *Trends in Biochemical Sciences,* 19**,** 114-118.

LLANO, M., SAENZ, D. T., MEEHAN, A., WONGTHIDA, P., PERETZ, M., WALKER, W. H., TEO, W. & POESCHLA, E. M. 2006. An Essential Role for LEDGF/p75 in HIV Integration. *Science,* 314**,** 461.

LLANO, M., VANEGAS, M., FREGOSO, O., SAENZ, D., CHUNG, S., PERETZ, M. & POESCHLA, E. M. 2004. LEDGF/p75 Determines Cellular Trafficking of Diverse Lentiviral but Not Murine Oncoretroviral Integrase Proteins and Is a Component of Functional Lentiviral Preintegration Complexes. *Journal of Virology,* 78**,** 9524.

LOCHS, S. J. A., KEFALOPOULOU, S. & KIND, J. 2019. Lamina Associated Domains and Gene Regulation in Development and Cancer. *Cells,* 8**,** 271.

LODI, P. J., ERNST, J. A., KUSZEWSKI, J., HICKMAN, A. B., ENGELMAN, A., CRAIGIE, R., CLORE, G. M. & GRONENBORN, A. M. 1995. Solution Structure of the DNA Binding Domain of HIV-1 Integrase. *Biochemistry,* 34**,** 9826-9833.

LU, R., GHORY, H. Z. & ENGELMAN, A. 2005a. Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase. *Journal of virology,* 79**,** 10356-10368.

LU, R., LIMÓN, A., GHORY, H. Z. & ENGELMAN, A. 2005b. Genetic Analyses of DNA-Binding Mutants in the Catalytic Core Domain of Human Immunodeficiency Virus Type 1 Integrase. *Journal of Virology,* 79**,** 2493.

LU, R., VANDEGRAAFF, N., CHEREPANOV, P. & ENGELMAN, A. 2005c. Lys-34, dispensable for integrase catalysis, is required for preintegration complex function and human immunodeficiency virus type 1 replication. *J Virol,* 79**,** 12584-91.

LUO, G. X., SHARMEEN, L. & TAYLOR, J. 1990. Specificities involved in the initiation of retroviral plus-strand DNA. *Journal of Virology,* 64**,** 592.

LUTZKE, R. A. & PLASTERK, R. H. 1998. Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. *Journal of virology,* 72**,** 4841-4848.

MADDON, P. J., DALGLEISH, A. G., MCDOUGAL, J. S., CLAPHAM, P. R., WEISS, R. A. & AXEL, R. 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell,* 47**,** 333-348.

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**,** 33528-39.

MAERTENS, G. N. 2016. B′ -protein phosphatase 2A is a functional binding partner of delta-retroviral integrase. *Nucleic Acids Research,* 44**,** 364-376.

MAERTENS, G. N., HARE, S. & CHEREPANOV, P. 2010. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature,* 468**,** 326-9.

MALLERY, D. L., MÁRQUEZ, C. L., MCEWAN, W. A., DICKSON, C. F., JACQUES, D. A., ANANDAPADAMANABAN, M., BICHEL, K., TOWERS, G. J., SAIARDI, A., BÖCKING, T. & JAMES, L. C. 2018. IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *eLife,* 7**,** e35335.

MARIN, M., ROSE, K. M., KOZAK, S. L. & KABAT, D. 2003. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nature Medicine,* 9**,** 1398-1403.

MARINI, B., KERTESZ-FARKAS, A., ALI, H., LUCIC, B., LISEK, K., MANGANARO, L., PONGOR, S., LUZZATI, R., RECCHIA, A., MAVILIO, F., GIACCA, M. & LUSIC, M. 2015. Nuclear architecture dictates HIV-1 integration site selection. *Nature,* 521**,** 227-231.

MARQUET, R., ISEL, C., EHRESMANN, C. & EHRESMANN, B. 1995. tRNAs as primer of reverse transcriptases. *Biochimie,* 77**,** 113-124.

MARTIN, G., GRUBER, ANDREAS R., KELLER, W. & ZAVOLAN, M. 2012. Genome-wide Analysis of Pre-mRNA 3′ End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3′ UTR Length. *Cell Rep,* 1**,** 753-763.

MARTIN-SERRANO, J., ZANG, T. & BIENIASZ, P. D. 2001. HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nature Medicine,* 7**,** 1313-1319.

MASKELL, D. P., RENAULT, L., SERRAO, E., LESBATS, P., MATADEEN, R., HARE, S., LINDEMANN, D., ENGELMAN, A. N., COSTA, A. & CHEREPANOV, P. 2015. Structural basis for retroviral integration into nucleosomes. *Nature,* 523**,** 366-9.

MATTEI, S., GLASS, B., HAGEN, W. J. H., KRÄUSSLICH, H.-G. & BRIGGS, J. A. G. 2016a. The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science,* 354**,** 1434.

MATTEI, S., SCHUR, F. K. M. & BRIGGS, J. A. G. 2016b. Retrovirus maturation—an extraordinary structural transformation. *Current Opinion in Virology,* 18**,** 27-35.

MCKEE, C. J., KESSL, J. J., SHKRIABAI, N., DAR, M. J., ENGELMAN, A. & KVARATSKHELIA, M. 2008. Dynamic modulation of HIV-1 integrase structure and function by cellular lens epithelium-derived growth factor (LEDGF) protein. *The Journal of biological chemistry,* 283**,** 31802-31812.

MCLEAN, R. K. & GRIFFITHS, D. J. Manuscript in preparation.

MEIERING, C. D. & LINIAL, M. L. 2001. Historical perspective of foamy virus epidemiology and infection. *Clin Microbiol Rev,* 14**,** 165-76.

MELAMED, A., LAYDON, D. J., GILLET, N. A., TANAKA, Y., TAYLOR, G. P. & BANGHAM, C. R. M. 2013. Genome-wide Determinants of Proviral Targeting, Clonal Abundance and Expression in Natural HTLV-1 Infection. *PLOS Pathogens,* 9**,** e1003271.

MILLER, M. D., FARNET, C. M. & BUSHMAN, F. D. 1997. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *Journal of Virology,* 71**,** 5382.

MITCHELL, R. S., BEITZEL, B. F., SCHRODER, A. R., SHINN, P., CHEN, H., BERRY, C. C., ECKER, J. R. & BUSHMAN, F. D. 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol,* 2**,** E234.

MOEBES, A., ENSSLE, J., BIENIASZ, P. D., HEINKELEIN, M., LINDEMANN, D., BOCK, M., MCCLURE, M. O. & RETHWILM, A. 1997. Human foamy virus reverse transcription that occurs late in the viral replication cycle. *J Virol,* 71**,** 7305-11.

MÜLLER, H. P. & VARMUS, H. E. 1994. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *The EMBO journal,* 13**,** 4704-4714.

NALAM, M. N. L., ALI, A., ALTMAN, M. D., REDDY, G. S. K. K., CHELLAPPAN, S., KAIRYS, V., ÖZEN, A., CAO, H., GILSON, M. K., TIDOR, B., RANA, T. M. & SCHIFFER, C. A. 2010. Evaluating the Substrate-Envelope Hypothesis: Structural Analysis of Novel HIV-1 Protease Inhibitors Designed To Be Robust against Drug Resistance. *Journal of Virology,* 84**,** 5368.

NARAYAN, O., WOLINSKY, J. S., CLEMENTS, J. E., STRANDBERG, J. D., GRIFFIN, D. E. & CORK, L. C. 1982. Slow Virus Replication: the Role of Macrophages in the Persistence and Expression of Visna Viruses of Sheep and Goats. *Journal of General Virology,* 59**,** 345-356.

NEIL, S. J. D., ZANG, T. & BIENIASZ, P. D. 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature,* 451**,** 425-430.

NISOLE, S. & SAÏB, A. 2004. Early steps of retrovirus replicative cycle. *Retrovirology,* 1**,** 9-9.

NITTA, T., KUZNETSOV, Y., MCPHERSON, A. & FAN, H. 2010. Murine leukemia virus glycosylated Gag (gPr80*gag*) facilitates interferon-sensitive virus release through lipid rafts. *Proceedings of the National Academy of Sciences,* 107**,** 1190.

NOMAGUCHI, M., MIYAKE, A., DOI, N., FUJIWARA, S., MIYAZAKI, Y., TSUNETSUGU-YOKOTA, Y., YOKOYAMA, M., SATO, H., MASUDA, T. & ADACHI, A. 2014. Natural Single-Nucleotide Polymorphisms in the 3′ Region of the HIV-1 *pol* Gene Modulate Viral Replication Ability. *Journal of Virology,* 88**,** 4145.

OLECH, M., MURAWSKI, M. & KUŹMAK, J. 2019. Molecular analysis of small-ruminant lentiviruses in Polish flocks reveals the existence of a novel subtype in sheep. *Archives of Virology,* 164**,** 1193-1198.

OLECH, M., RACHID, A., CROISÉ, B., KUŹMAK, J. & VALAS, S. 2012. Genetic and antigenic characterization of small ruminant lentiviruses circulating in Poland. *Virus Research,* 163**,** 528-536.

OSAME, M., USUKU, K., IZUMO, S., IJICHI, N., AMITANI, H., IGATA, A., MATSUMOTO, M. & TARA, M. 1986. HTLV-I Associated Myelopathy, A New Clinical Entity. *The Lancet,* 327**,** 1031-1032.

OTWINOWSKI, Z., SCHEVITZ, R. W., ZHANG, R. G., LAWSON, C. L., JOACHIMIAK, A., MARMORSTEIN, R. Q., LUISI, B. F. & SIGLER, P. B. 1988. Crystal structure of trp represser/operator complex at atomic resolution. *Nature,* 335**,** 321-329.

OZOROWSKI, G., PALLESEN, J., DE VAL, N., LYUMKIS, D., COTTRELL, C. A., TORRES, J. L., COPPS, J., STANFIELD, R. L., CUPO, A., PUGACH, P., MOORE, J. P., WILSON, I. A. & WARD, A. B. 2017. Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike. *Nature,* 547**,** 360-363.

PAIVA, A. & CASSEB, J. 2015. Origin and prevalence of human T-lymphotropic virus type 1 (HTLV-1) and type 2 (HTLV-2) among indigenous populations in the Americas. *Rev Inst Med Trop Sao Paulo,* 57**,** 1-13.

PANCERA, M., MAJEED, S., BAN, Y.-E. A., CHEN, L., HUANG, C.-C., KONG, L., KWON, Y. D., STUCKEY, J., ZHOU, T., ROBINSON, J. E., SCHIEF, W. R., SODROSKI, J., WYATT, R. & KWONG, P. D. 2010. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proceedings of the National Academy of Sciences,* 107**,** 1166.

PANDEY, K. K., BERA, S., SHI, K., RAU, M. J., OLERU, A. V., FITZPATRICK, J. A. J., ENGELMAN, A. N., AIHARA, H. & GRANDGENETT, D. P. 2021. Cryo-EM structure of the Rous sarcoma virus octameric cleaved synaptic complex intasome. *Communications Biology,* 4**,** 330.

PANDEY, K. K., BERA, S., ZAHM, J., VORA, A., STILLMOCK, K., HAZUDA, D. & GRANDGENETT, D. P. 2007. Inhibition of human immunodeficiency virus type 1 concerted integration by strand transfer inhibitors which recognize a transient structural intermediate. *Journal of virology,* 81**,** 12189-12199.

PARHAM, J. H., IANNONE, M. A., OVERTON, L. K. & HUTCHINS, J. T. 1998. Optimization of transient gene expression in mammalian cells and potential for scale-up using flow electroporation. *In:* BETENBAUGH, M. J., CHALMERS, J. J., ARATHOON, R., CHAPLEN, F. W. R. & MASTRANGELO, A. J. (eds.) *Cell Culture Engineering VI.* Dordrecht: Springer Netherlands.

PASSOS, D. O., LI, M., JÓŹWIK, I. K., ZHAO, X. Z., SANTOS-MARTINS, D., YANG, R., SMITH, S. J., JEON, Y., FORLI, S., HUGHES, S. H., BURKE, T. R., CRAIGIE, R. & LYUMKIS, D. 2020. Structural basis for strand-transfer inhibitor binding to HIV intasomes. *Science,* 367**,** 810.

PASSOS, D. O., LI, M., YANG, R., REBENSBURG, S. V., GHIRLANDO, R., JEON, Y., SHKRIABAI, N., KVARATSKHELIA, M., CRAIGIE, R. & LYUMKIS, D. 2017. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science,* 355**,** 89.

PEDERSEN, N. C., HO, E. W., BROWN, M. L. & YAMAMOTO, J. K. 1987. Isolation of a T-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome. *Science,* 235**,** 790.

PEETERS, M., HONORÉ, C., HUET, T., BEDJABAGA, L., OSSARI, S., BUSSI, P., COOPER, R. W. & DELAPORTE, E. 1989. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS,* 3.

PELISKA, J. A. & BENKOVIC, S. J. 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science,* 258**,** 1112.

PEPIN, M., VITU, C., RUSSO, P., MORNEX, J. F. & PETERHANS, E. 1998. Maedi-visna virus infection in sheep: a review. *Vet Res,* 29**,** 341-67.

PETTIT, S. C., MOODY, M. D., WEHBIE, R. S., KAPLAN, A. H., NANTERMET, P. V., KLEIN, C. A. & SWANSTROM, R. 1994. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *Journal of virology,* 68**,** 8017-8027.

PHONGSAVANH, X., AL-QATABI, N., SHABAN, M. S., KHODER-AGHA, F., EL ASRI, M., COMISSO, M., GUÉROIS, R. & MIRANDE, M. 2020. How HIV-1 Integrase Associates with Human Mitochondrial Lysyl-tRNA Synthetase. *Viruses,* 12.

PIENIAZEK, D., ELLENBERGER, D., JANINI, L. M., RAMOS, A. C., NKENGASONG, J., SASSAN-MOROKRO, M., HU, D. J., COULIBALLY, I.-M., EKPINI, E., BANDEA, C., TANURI, A., GREENBERG, A. E., WIKTOR, S. Z. & RAYFIELD, M. A. 1999. Predominance of Human Immunodeficiency Virus Type 2 Subtype B in Abidjan, Ivory Coast. *AIDS Research and Human Retroviruses,* 15**,** 603-608.

PLANTIER, J.-C., LEOZ, M., DICKERSON, J. E., DE OLIVEIRA, F., CORDONNIER, F., LEMÉE, V., DAMOND, F., ROBERTSON, D. L. & SIMON, F. 2009. A new human immunodeficiency virus derived from gorillas. *Nature Medicine,* 15**,** 871-872.

PLUYMERS, W., CHEREPANOV, P., SCHOLS, D., DE CLERCQ, E. & DEBYSER, Z. 1999. Nuclear localization of human immunodeficiency virus type 1 integrase expressed as a fusion protein with green fluorescent protein. *Virology,* 258**,** 327-32.

POIESZ, B. J., RUSCETTI, F. W., GAZDAR, A. F., BUNN, P. A., MINNA, J. D. & GALLO, R. C. 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proceedings of the National Academy of Sciences,* 77**,** 7415.

POPOVIC, M., SARNGADHARAN, M. G., READ, E. & GALLO, R. C. 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science,* 224**,** 497.

POPPER, S. J., SARR, A. D., GUÈYE-NDIAYE, A., MBOUP, S., ESSEX, M. E. & KANKI, P. J. 2000. Low Plasma Human Immunodeficiency Virus Type 2 Viral Load Is Independent of Proviral Load: Low Virus Production In Vivo. *Journal of Virology,* 74**,** 1554.

PRABU-JEYABALAN, M., NALIVAIKA, E. & SCHIFFER, C. A. 2002. Substrate Shape Determines Specificity of Recognition for HIV-1 Protease: Analysis of Crystal Structures of Six Substrate Complexes. *Structure,* 10**,** 369-381.

PRUSS, D., BUSHMAN, F. D. & WOLFFE, A. P. 1994. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proceedings of the National Academy of Sciences,* 91**,** 5913.

PRYCIAK, P. M. & VARMUS, H. E. 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell,* 69**,** 769-80.

PURCELL, D. F. & MARTIN, M. A. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *Journal of Virology,* 67**,** 6365.

QUERAT, G., AUDOLY, G., SONIGO, P. & VIGNE, R. 1990. Nucleotide sequence analysis of SA-OMVV, a visna-related ovine lentivirus: phylogenetic history of lentiviruses. *Virology,* 175**,** 434-447.

RAGHAVENDRA, N. K., SHKRIABAI, N., GRAHAM, R. L. J., HESS, S., KVARATSKHELIA, M. & WU, L. 2010. Identification of host proteins associated with HIV-1 preintegration complexes isolated from infected CD4+ cells. *Retrovirology,* 7**,** 66.

RAI, M. A., PANNEK, S. & FICHTENBAUM, C. J. 2018. Emerging reverse transcriptase inhibitors for HIV-1 infection. *Expert Opin Emerg Drugs,* 23**,** 149-157.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols,* 8**,** 2281-2308.

RASAIYAAH, J., TAN, C. P., FLETCHER, A. J., PRICE, A. J., BLONDEAU, C., HILDITCH, L., JACQUES, D. A., SELWOOD, D. L., JAMES, L. C., NOURSADEGHI, M. & TOWERS, G. J. 2013. HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature,* 503**,** 402-405.

RATNER, L., HASELTINE, W., PATARCA, R., LIVAK, K. J., STARCICH, B., JOSEPHS, S. F., DORAN, E. R., RAFALSKI, J. A., WHITEHORN, E. A., BAUMEISTER, K., IVANOFF, L., PETTEWAY, S. R., PEARSON, M. L., LAUTENBERGER, J. A., PAPAS, T. S., GHRAYEB, J., CHANG, N. T., GALLO, R. C. & WONG-STAAL, F. 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature,* 313**,** 277-284.

REINA, R., MORA, M. I., GLARIA, I., GARCÍA, I., SOLANO, C., LUJÁN, L., BADIOLA, J. J., CONTRERAS, A., BERRIATUA, E., JUSTE, R., MAMOUN, R. Z., ROLLAND, M., AMORENA, B. & DE ANDRÉS, D. 2006. Molecular characterization and phylogenetic study of Maedi Visna and Caprine Arthritis Encephalitis viral sequences in sheep and goats from Spain. *Virus Research,* 121**,** 189-198.

REYNOLDS, E. S. 1963. The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. *J Cell Biol,* 17**,** 208-212.

ROBERTSON, D. L., ANDERSON, J. P., BRADAC, J. A., CARR, J. K., FOLEY, B., FUNKHOUSER, R. K., GAO, F., HAHN, B. H., KALISH, M. L., KUIKEN, C., LEARN, G. H., LEITNER, T., MCCUTCHAN, F., OSMANOV, S., PEETERS, M., PIENIAZEK, D., SALMINEN, M., SHARP, P. M., WOLINSKY, S. & KORBER, B. 2000. HIV-1 Nomenclature Proposal. *Science,* 288**,** 55.

ROCA, A. I. & COX, M. M. 1997. RecA Protein: Structure, Function, and Role in Recombinational DNA Repair. *In:* COHN, W. E. & MOLDAVE, K. (eds.) *Progress in Nucleic Acid Research and Molecular Biology.* Academic Press.

ROMANI, B., SHAYKH BAYGLOO, N., AGHASADEGHI, M. R. & ALLAHBAKHSHI, E. 2015. HIV-1 Vpr Protein Enhances Proteasomal Degradation of MCM10 DNA Replication Factor through the Cul4-DDB1[VprBP] E3 Ubiquitin Ligase to Induce G2/M Cell Cycle Arrest*. *Journal of Biological Chemistry,* 290**,** 17380-17389.

ROSA, A., CHANDE, A., ZIGLIO, S., DE SANCTIS, V., BERTORELLI, R., GOH, S. L., MCCAULEY, S. M., NOWOSIELSKA, A., ANTONARAKIS, S. E., LUBAN, J., SANTONI, F. A. & PIZZATO, M. 2015. HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature,* 526**,** 212-217.

ROSS, T. M., ORAN, A. E. & CULLEN, B. R. 1999. Inhibition of HIV-1 progeny virion release by cell-surface CD4 is relieved by expression of the viral Nef protein. *Current Biology,* 9**,** 613-621.

ROTH, S. L., MALANI, N. & BUSHMAN, F. D. 2011. Gammaretroviral integration into nucleosomal target DNA in vivo. *Journal of virology,* 85**,** 7393-7401.

RÜEGSEGGER, U., BEYER, K. & KELLER, W. 1996. Purification and Characterization of Human Cleavage Factor Im Involved in the 3′ End Processing of Messenger RNA Precursors. *Journal of Biological Chemistry,* 271**,** 6107-6113.

SALTARELLI, M., QUERAT, G., KONINGS, D. A. M., VIGNE, R. & CLEMENTS, J. E. 1990. Nucleotide sequence and transcriptional analysis of molecular clones of CAEV which generate infectious virus. *Virology,* 179**,** 347-364.

SARAFIANOS, S. G., MARCHAND, B., DAS, K., HIMMEL, D. M., PARNIAK, M. A., HUGHES, S. H. & ARNOLD, E. 2009. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol,* 385**,** 693-713.

SARGAN, D. R., BENNET, I. D., COUSENS, C., ROY, D. J., BLACKLAWS, B. A., DALZIEL, R. G., WATT, N. J. & MCCONNELL, I. 1991. Nucleotide Sequence of EV1, a British Isolate Of Maedi-Visna Virus. *Journal of General Virology,* 72**,** 1893-1903.

SCHALLER, T., OCWIEJA, K. E., RASAIYAAH, J., PRICE, A. J., BRADY, T. L., ROTH, S. L., HUÉ, S., FLETCHER, A. J., LEE, K., KEWALRAMANI, V. N., NOURSADEGHI, M., JENNER, R. G., JAMES, L. C., BUSHMAN, F. D. & TOWERS, G. J. 2011. HIV-1 Capsid-Cyclophilin Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency. *PLOS Pathogens,* 7**,** e1002439.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J.-Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods,* 9**,** 676-682.

SCHIPPER, I., MISEK, A., LUDEMANN, L., LIGHT, M. & LIMESAND, W. 1983. Ovine progressive pneumonia infection via the oral route. *VETERINARY MEDICINE & SMALL ANIMAL CLINICIAN,* 78**,** 415-417.

SCHRIJVERS, R., DE RIJCK, J., DEMEULEMEESTER, J., ADACHI, N., VETS, S., RONEN, K., CHRIST, F., BUSHMAN, F. D., DEBYSER, Z. & GIJSBERS, R. 2012a. LEDGF/p75-independent HIV-1 replication demonstrates a role for HRP-2 and remains sensitive to inhibition by LEDGINs. *PLoS pathogens,* 8**,** e1002558-e1002558.

SCHRIJVERS, R., VETS, S., DE RIJCK, J., MALANI, N., BUSHMAN, F. D., DEBYSER, Z. & GIJSBERS, R. 2012b. HRP-2 determines HIV-1 integration site selection in LEDGF/p75 depleted cells. *Retrovirology,* 9**,** 84-84.

SCHRODER, A. R., SHINN, P., CHEN, H., BERRY, C., ECKER, J. R. & BUSHMAN, F. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell,* 110**,** 521-9.

SCHRÖDER, A. R. W., SHINN, P., CHEN, H., BERRY, C., ECKER, J. R. & BUSHMAN, F. 2002. HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell,* 110**,** 521-529.

SCHWARTZ, O., MARÉCHAL, V., GALL, S. L., LEMONNIER, F. & HEARD, J.-M. 1996. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV–1 Nef protein. *Nature Medicine,* 2**,** 338-342.

SCHWEITZER, C. J., JAGADISH, T., HAVERLAND, N., CIBOROWSKI, P. & BELSHAN, M. 2013. Proteomic analysis of early HIV-1 nucleoprotein complexes. *J Proteome Res,* 12**,** 559-572.

SERRAO, E., BALLANDRAS-COLAS, A., CHEREPANOV, P., MAERTENS, G. N. & ENGELMAN, A. N. 2015. Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology,* 12**,** 39.

SERRAO, E., CHEREPANOV, P. & ENGELMAN, A. N. 2016. Amplification, Next-generation Sequencing, and Genomic DNA Mapping of Retroviral Integration Sites. *J Vis Exp*.

SERRAO, E., KRISHNAN, L., SHUN, M. C., LI, X., CHEREPANOV, P., ENGELMAN, A. & MAERTENS, G. N. 2014. Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: implications for the mechanism of target DNA binding. *Nucleic Acids Res,* 42**,** 5164-76.

SHAH, C., BÖNI, J., HUDER, J. B., VOGT, H.-R., MÜHLHERR, J., ZANONI, R., MISEREZ, R., LUTZ, H. & SCHÜPBACH, J. 2004. Phylogenetic analysis and reclassification of caprine and ovine lentiviruses based on 104 new isolates: evidence for regular sheep-to-goat transmission and worldwide propagation through livestock trade. *Virology,* 319**,** 12-26.

SHARMA, A., LARUE, R. C., PLUMB, M. R., MALANI, N., MALE, F., SLAUGHTER, A., KESSL, J. J., SHKRIABAI, N., COWARD, E., AIYER, S. S., GREEN, P. L., WU, L., ROTH, M. J., BUSHMAN, F. D. & KVARATSKHELIA, M. 2013. BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 12036-12041.

SHARP, P. M. & HAHN, B. H. 2011. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med,* 1**,** a006841-a006841.

SHAW, G. M., HAHN, B. H., ARYA, S. K., GROOPMAN, J. E., GALLO, R. C. & WONG-STAAL, F. 1984. Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. *Science,* 226**,** 1165.

SHIN, C. G., TADDEO, B., HASELTINE, W. A. & FARNET, C. M. 1994. Genetic analysis of the human immunodeficiency virus type 1 integrase protein. *Journal of virology,* 68**,** 1633-1642.

SHUN, M. C., RAGHAVENDRA, N. K., VANDEGRAAFF, N., DAIGLE, J. E., 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**,** 1767-78.

SIDDIQUI, M. A., SAITO, A., HALAMBAGE, U. D., FERHADIAN, D., FISCHER, D. K., FRANCIS, A. C., MELIKYAN, G. B., AMBROSE, Z., AIKEN, C. & YAMASHITA, M. 2019. A Novel Phenotype Links HIV-1 Capsid Stability to cGAS-Mediated DNA Sensing. *Journal of Virology,* 93**,** e00706-19.

SIGURDSSON, B. 1954. Mædi, A Slow Progressive Pneumonia of Sheep: An Epizoological and a Pathological Study. *British Veterinary Journal,* 110**,** 255-270.

SIHVONEN, L., HIRVELÄ-KOSKI, V., NUOTIO, L. & KOKKONEN, U. M. 1999. Serological survey and epidemiological investigation of maedi-visna in sheep in Finland. *Veterinary Microbiology,* 65**,** 265-270.

SILVA, A. M., CACHAU, R. E., SHAM, H. L. & ERICKSON, J. W. 1996. Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J Mol Biol,* 255**,** 321-340.

SINGH, P. K., PLUMB, M. R., FERRIS, A. L., IBEN, J. R., WU, X., FADEL, H. J., LUKE, B. T., ESNAULT, C., POESCHLA, E. M., HUGHES, S. H., KVARATSKHELIA, M. & LEVIN, H. L. 2015. LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. *Genes & Development,* 29**,** 2287-2297.

SKOKO, D., LI, M., HUANG, Y., MIZUUCHI, M., CAI, M., BRADLEY, C. M., PEASE, P. J., XIAO, B., MARKO, J. F., CRAIGIE, R. & MIZUUCHI, K. 2009. Barrier-to-autointegration factor (BAF) condenses DNA by looping. *Proceedings of the National Academy of Sciences,* 106**,** 16610.

SODROSKI, J., GOH, W. C., ROSEN, C., DAYTON, A., TERWILLIGER, E. & HASELTINE, W. 1986. A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature,* 321**,** 412-417.

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**,** 369-382.

SOWD, G. A., SERRAO, E., WANG, H., WANG, W., FADEL, H. J., POESCHLA, E. M. & ENGELMAN, A. N. 2016. A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proceedings of the National Academy of Sciences,* 113**,** E1054-E1063.

SPENCE, R. A., KATI, W. M., ANDERSON, K. S. & JOHNSON, K. A. 1995. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Science,* 267**,** 988.

STEC, I., NAGL, S. B., VAN OMMEN, G.-J. B. & DEN DUNNEN, J. T. 2000. The PWWP domain: a potential protein–protein interaction domain in nuclear proteins influencing differentiation? *FEBS Letters,* 473**,** 1-5.

STEIN, B. S., GOWDA, S. D., LIFSON, J. D., PENHALLOW, R. C., BENSCH, K. G. & ENGLEMAN, E. G. 1987. pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. *Cell,* 49**,** 659-668.

STEVENS, S. W. & GRIFFITH, J. D. 1996. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. *Journal of virology,* 70**,** 6459-6462.

STOPAK, K., DE NORONHA, C., YONEMOTO, W. & GREENE, W. C. 2003. HIV-1 Vif Blocks the Antiviral Activity of APOBEC3G by Impairing Both Its Translation and Intracellular Stability. *Molecular Cell,* 12**,** 591-601.

STUDIER, F. W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol,* 219**,** 37-44.

STUMPTNER-CUVELETTE, P., MORCHOISNE, S., DUGAST, M., LE GALL, S., RAPOSO, G., SCHWARTZ, O. & BENAROCH, P. 2001. HIV-1 Nef impairs MHC class II antigen presentation and surface expression. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 12144-12149.

SUMNER, R. P., HARRISON, L., TOUIZER, E., PEACOCK, T. P., SPENCER, M., ZULIANI-ALVAREZ, L. & TOWERS, G. J. 2020. Disrupting HIV-1 capsid formation causes cGAS sensing of viral DNA. *The EMBO Journal,* 39**,** e103958.

SUVEGES, T., SUVEGES, T. & SZEKY, A. 1973. Incidence of Maedi (chronic progressive interstitial pneumonia) among sheep in Hungary. *Acta Vet Acad Sci Hung,* 23**,** 205-17.

TADDEO, B., HASELTINE, W. A. & FARNET, C. M. 1994. Integrase mutants of human immunodeficiency virus type 1 with a specific defect in integration. *Journal of Virology,* 68**,** 8401-8405.

TAYLOR, R. G., WALKER, D. C. & MCLNNES, R. R. 1993. *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research,* 21**,** 1677-1678.

TEBIT, D. M. & ARTS, E. J. 2011. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *The Lancet Infectious Diseases,* 11**,** 45-56.

TEKESTE, S. S., WILKINSON, T. A., WEINER, E. M., XU, X., MILLER, J. T., LE GRICE, S. F. J., CLUBB, R. T. & CHOW, S. A. 2015. Interaction between Reverse Transcriptase and Integrase Is Required for Reverse Transcription during HIV-1 Replication. *Journal of virology,* 89**,** 12058-12069.

TEMIN, H. M. & MIZUTANI, S. 1970. Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. *Nature,* 226**,** 1211-1213.

THORMAR, H., BALZARINI, J., HOLY, A., JINDRICH, J., ROSENBERG, I., DEBYSER, Z., DESMYTER, J. & DE CLERCQ, E. 1993. Inhibition of visna virus replication by 2',3'-dideoxynucleosides and acyclic nucleoside phosphonate analogs. *Antimicrobial agents and chemotherapy,* 37**,** 2540-2544.

THORMAR, H. & CRUICKSHANK, J. G. 1965. The Structure of Visna Virus Studied by the Negative Staining Technique. *Virology,* 25**,** 145-8.

TRKOLA, A., DRAGIC, T., ARTHOS, J., BINLEY, J. M., OLSON, W. C., ALLAWAY, G. P., CHENG-MAYER, C., ROBINSON, J., MADDON, P. J. & MOORE, J. P. 1996. CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature,* 384**,** 184-187.

TSIANG, M., JONES, G. S., NIEDZIELA-MAJKA, A., KAN, E., LANSDON, E. B., HUANG, W., HUNG, M., SAMUEL, D., NOVIKOV, N., XU, Y., MITCHELL, M., GUO, H., BABAOGLU, K., LIU, X., GELEZIUNAS, R. & SAKOWICZ, R. 2012. New Class of HIV-1 Integrase (IN) Inhibitors with a Dual Mode of Action. *Journal of Biological Chemistry,* 287**,** 21189-21203.

TSOUROULA, K., FURST, A., ROGIER, M., HEYER, V., MAGLOTT-ROTH, A., FERRAND, A., REINA-SAN-MARTIN, B. & SOUTOGLOU, E. 2016. Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Molecular Cell,* 63**,** 293-305.

TURELLI, P., PÉTURSSON, G., GUIGUEN, F., MORNEX, J. F., VIGNE, R. & QUÉRAT, G. 1996. Replication properties of dUTPase-deficient mutants of caprine and ovine lentiviruses. *Journal of Virology,* 70**,** 1213.

TURLURE, F., MAERTENS, G., RAHMAN, S., CHEREPANOV, P. & ENGELMAN, A. 2006. A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. *Nucleic Acids Research,* 34**,** 1653-1665.

ULM, J. W., PERRON, M., SODROSKI, J. & C. MULLIGAN, R. 2007. Complex determinants within the Moloney murine leukemia virus capsid modulate susceptibility of the virus to Fv1 and Ref1-mediated restriction. *Virology,* 363**,** 245-255.

UNAIDS 2020. Fact Sheet and Global HIV statistics.

VALKOV, E., GUPTA, S. S., HARE, S., HELANDER, A., ROVERSI, P., MCCLURE, M. & CHEREPANOV, P. 2009. Functional and structural characterization of the integrase from the prototype foamy virus. *Nucleic Acids Res,* 37**,** 243-55.

VALLARI, A., HOLZMAYER, V., HARRIS, B., YAMAGUCHI, J., NGANSOP, C., MAKAMCHE, F., MBANYA, D., KAPTUÉ, L., NDEMBI, N., GÜRTLER, L., DEVARE, S. & BRENNAN, C. A. 2011. Confirmation of Putative HIV-1 Group P in Cameroon. *Journal of Virology,* 85**,** 1403.

VAN DER MAATEN, M. J., BOOTHE, A. D. & SEGER, C. L. 1972. Isolation of a Virus From Cattle With Persistent Lymphocytosis. *JNCI: Journal of the National Cancer Institute,* 49**,** 1649-1657.

VAN HEUVERSWYN, F., LI, Y., NEEL, C., BAILES, E., KEELE, B. F., LIU, W., LOUL, S., BUTEL, C., LIEGEOIS, F., BIENVENUE, Y., NGOLLE, E. M., SHARP, P. M., SHAW, G. M., DELAPORTE, E., HAHN, B. H. & PEETERS, M. 2006. SIV infection in wild gorillas. *Nature,* 444**,** 164-164.

VANDEGRAAFF, N., DEVROE, E., TURLURE, F., SILVER, P. A. & ENGELMAN, A. 2006. Biochemical and genetic analyses of integrase-interacting proteins lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication. *Virology,* 346**,** 415-426.

VANEGAS, M., LLANO, M., DELGADO, S., THOMPSON, D., PERETZ, M. & POESCHLA, E. 2005. Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. *Journal of Cell Science,* 118**,** 1733.

VANIN, E. F. 1985. Processed preudogenes: Characteristics and Evolution. *Annual Review of Genetics,* 19**,** 253-272.

VERMEIRE, J., NAESSENS, E., VANDERSTRAETEN, H., LANDI, A., IANNUCCI, V., VAN NUFFEL, A., TAGHON, T., PIZZATO, M. & VERHASSELT, B. 2012. Quantification of Reverse Transcriptase Activity by Real-Time PCR as a Fast and Accurate Method for Titration of HIV, Lenti- and Retroviral Vectors. *PLOS ONE,* 7**,** e50859.

VIGNE, R., FILIPPI, P., QUÉRAT, G., SAUZE, N., VITU, C., RUSSO, P. & DELORI, P. 1982. Precursor Polypeptides to Structural Proteins of Visna Virus. *Journal of Virology,* 42**,** 1046.

VILLET, S., BOUZAR, B. A., MORIN, T., VERDIER, G., LEGRAS, C. & CHEBLOUNE, Y. 2003. Maedi-visna virus and caprine arthritis encephalitis virus genomes encode a Vpr-like but no Tat protein. *J Virol,* 77**,** 9632-8.

WANG, G. P., CIUFFI, A., LEIPZIG, J., BERRY, C. C. & BUSHMAN, F. D. 2007. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res,* 17**,** 1186-1194.

WANG, H., JURADO, K. A., WU, X., SHUN, M. C., LI, X., FERRIS, A. L., SMITH, S. J., PATEL, P. A., FUCHS, J. R., CHEREPANOV, P., KVARATSKHELIA, M., HUGHES, S. H. & ENGELMAN, A. 2012. HRP2 determines the efficiency and specificity of HIV-1 integration in LEDGF/p75 knockout cells but does not contribute to the antiviral activity of a potent LEDGF/p75-binding site integrase inhibitor. *Nucleic Acids Res,* 40**,** 11518-30.

WANG, J. Y., LING, H., YANG, W. & CRAIGIE, R. 2001. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J,* 20**,** 7333-43.

WANG, S.-W. & ALDOVINI, A. 2002. RNA Incorporation Is Critical for Retroviral Particle Integrity after Cell Membrane Assembly of Gag Complexes. *Journal of Virology,* 76**,** 11853.

WANG, X., SINGH, S., JUNG, H.-Y., YANG, G., JUN, S., SASTRY, K. J. & PARK, J.-I. 2013. HIV-1 Vpr Protein Inhibits Telomerase Activity via the EDD-DDB1-VPRBP E3 Ligase Complex*. *Journal of Biological Chemistry,* 288**,** 15474-15480.

WANG, Y., PROSEN, D. E., MEI, L., SULLIVAN, J. C., FINNEY, M. & VANDER HORN, P. B. 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic acids research,* 32**,** 1197-1207.

WEI, S. Q., MIZUUCHI, K. & CRAIGIE, R. 1997. A large nucleoprotein assembly at the ends of the viral DNA mediates retroviral DNA integration. *EMBO J,* 16**,** 7511-20.

WEILAND, F. & BRUNS, M. 1980. Ultrastructural studies on Maedi-Visna virus. *Arch Virol,* 64**,** 277-85.

WEISS, R. A. 1996. Retrovirus classification and cell interactions. *J Antimicrob Chemother,* 37 Suppl B**,** 1-11.

WHO. 2019. *Update of recommendations on first- and second-line antiretroviral regimens* [Online]. WHO Policy Brief, 1-16 (2019). Available: https://www.who.int/hiv/pub/arv/arv-update-2019-policy/en/ [Accessed].

WIEGERS, K., RUTTER, G., KOTTLER, H., TESSMER, U., HOHENBERG, H. & KRÄUSSLICH, H.-G. 1998. Sequential Steps in Human Immunodeficiency Virus

Particle Maturation Revealed by Alterations of Individual Gag Polyprotein Cleavage Sites. *Journal of Virology,* 72**,** 2846.

WIGHT, D. J., BOUCHERIT, V. C., WANAGURU, M., ELIS, E., HIRST, E. M. A., LI, W., EHRLICH, M., BACHARACH, E. & BISHOP, K. N. 2014. The N-Terminus of Murine Leukaemia Virus p12 Protein Is Required for Mature Core Stability. *PLOS Pathogens,* 10**,** e1004474.

WILKINSON, T. A., JANUSZYK, K., PHILLIPS, M. L., TEKESTE, S. S., ZHANG, M., MILLER, J. T., LE GRICE, S. F. J., CLUBB, R. T. & CHOW, S. A. 2009. Identifying and characterizing a functional HIV-1 reverse transcriptase-binding site on integrase. *The Journal of biological chemistry,* 284**,** 7931-7939.

WILLEY, R. L., MALDARELLI, F., MARTIN, M. A. & STREBEL, K. 1992. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. *Journal of Virology,* 66**,** 7193.

WILSON, M. D., RENAULT, L., MASKELL, D. P., GHONEIM, M., PYE, V. E., NANS, A., RUEDA, D. S., CHEREPANOV, P. & COSTA, A. 2019. Retroviral integration into nucleosomes through DNA looping and sliding along the histone octamer. *Nature Communications,* 10**,** 4189.

WISKERCHEN, M. & MUESING, M. A. 1995. Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. *Journal of Virology,* 69**,** 376-386.

WITVROUW, M., PANNECOUQUE, C., VAN LAETHEM, K., DESMYTER, J., DE CLERCQ, E. & VANDAMME, A.-M. 1999. Activity of non-nucleoside reverse transcriptase inhibitors against HIV-2 and SIV. *AIDS,* 13.

WLODAWER, A. & ERICKSON, J. W. 1993. STRUCTURE-BASED INHIBITORS OF HIV-1 PROTEASE. *Annual Review of Biochemistry,* 62**,** 543-585.

WU, X., LI, Y., CRISE, B. & BURGESS, S. M. 2003. Transcription Start Regions in the Human Genome Are Favored Targets for MLV Integration. *Science,* 300**,** 1749.

WU, X., LI, Y., CRISE, B., BURGESS, S. M. & MUNROE, D. J. 2005. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *Journal of virology,* 79**,** 5211-5214.

WU, X., LIU, H., XIAO, H., CONWAY, J. A., HEHL, E., KALPANA, G. V., PRASAD, V. & KAPPES, J. C. 1999. Human Immunodeficiency Virus Type 1 Integrase Protein Promotes Reverse Transcription through Specific Interactions with the Nucleoprotein Reverse Transcription Complex. *Journal of Virology,* 73**,** 2126-2135.

WYATT, R. & SODROSKI, J. 1998. The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens. *Science,* 280**,** 1884-1888.

YAMASHITA, M. & EMERMAN, M. 2004. Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *Journal of virology,* 78**,** 5670-5678.

YANG, K., GUO, R. & XU, D. 2016. Non-homologous end joining: advances and frontiers. *Acta Biochimica et Biophysica Sinica,* 48**,** 632-640.

YIN, Z., SHI, K., BANERJEE, S., PANDEY, K. K., BERA, S., GRANDGENETT, D. P. & AIHARA, H. 2016. Crystal structure of the Rous sarcoma virus intasome. *Nature,* 530**,** 362-366.

YODER, K., SARASIN, A., KRAEMER, K., MCILHATTON, M., BUSHMAN, F. & FISHEL, R. 2006. The DNA repair genes XPB and XPD defend cells from retroviral infection. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 4622-4627.

YODER, K. E. & BUSHMAN, F. D. 2000. Repair of gaps in retroviral DNA integration intermediates. *J Virol,* 74**,** 11191-200.

YU, S. F., SULLIVAN, M. D. & LINIAL, M. L. 1999. Evidence that the human foamy virus genome is DNA. *J Virol,* 73**,** 1565-72.

ZHAO, G., PERILLA, J. R., YUFENYUY, E. L., MENG, X., CHEN, B., NING, J., AHN, J., GRONENBORN, A. M., SCHULTEN, K., AIKEN, C. & ZHANG, P. 2013. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature,* 497**,** 643-646.

ZHAO, Z., MCKEE, C. J., KESSL, J. J., SANTOS, W. L., DAIGLE, J. E., ENGELMAN, A., VERDINE, G. & KVARATSKHELIA, M. 2008. Subunit-specific Protein Footprinting Reveals Significant Structural Rearrangements and a Role for N-terminal Lys-14 of HIV-1 Integrase during Viral DNA Binding. *Journal of Biological Chemistry,* 283**,** 5632-5641.

ZHENG, R., JENKINS, T. M. & CRAIGIE, R. 1996. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proceedings of the National Academy of Sciences,* 93**,** 13659-13664.

ZHOU, H.-X. & PANG, X. 2018. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chemical Reviews,* 118**,** 1691-1741.