



Etude fonctionnelle de l'interaction entre l'intasome du VIH-1 et le nucléosome : la queue d'histone H4 comme nouveau partenaire de l'intégration

Eric Mauro

► To cite this version:

Eric Mauro. Etude fonctionnelle de l'interaction entre l'intasome du VIH-1 et le nucléosome : la queue d'histone H4 comme nouveau partenaire de l'intégration. Médecine humaine et pathologie. Université de Bordeaux, 2018. Français. NNT : 2018BORD0273 . tel-02101175v2

HAL Id: tel-02101175

<https://theses.hal.science/tel-02101175v2>

Submitted on 16 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE SCIENCES DE LA VIE ET DE LA SANTÉ

SPÉCIALITÉ MICROBIOLOGIE – IMMUNOLOGIE

Par Eric MAURO

**Etude fonctionnelle de l'interaction entre l'intasome du VIH-1
et le nucléosome : la queue d'histone H4 comme nouveau
partenaire de l'intégration**

Sous la direction de : Vincent PARISSI

Soutenue le 3 Décembre 2018

Membres du jury :

Mr Thierry Noël, Professeur Universitaire, Bordeaux.....Président
Mme Francesca Di-Nunzio, Chargée de Recherche, Paris.....Rapporteur
Mr Matteo Negroni, Directeur de Recherche, Strasbourg.....Rapporteur
Mr Gael Cristofari, Directeur de Recherche, Nice.....Examineur
Mr Philippe Pourquier, Directeur de Recherche, Montpellier.....Examineur
Mr Vincent Parissi, Directeur de Recherche, Bordeaux.....Directeur de thèse

Etude fonctionnelle de l'interaction entre l'intasome du VIH-1 et le nucléosome : la queue d'histone H4 comme nouveau partenaire de l'intégration

Résumé : L'intégrase (IN) du VIH-1 est une enzyme qui catalyse l'intégration du génome du virus dans celui de la cellule infectée. Cette étape d'intégration est cruciale pour le virus pour qu'il puisse se répliquer de manière efficace, l'intégration est donc une cible de choix dans les thérapies antivirales. Comprendre les mécanismes qui participent à l'intégration est donc nécessaire afin de développer des solutions efficaces pour contrecarrer le virus.

L'intégration rétrovirale est catalysée par une structure oligomérique d'IN et d'ADN viral bien particulière appelée intasome. Les intasomes rétroviraux catalysent l'intégration préférentiellement sur des nucléosomes, composés d'ADN enroulé de protéines histones, plutôt que sur de l'ADN nu. Ceci est en partie dû aux contraintes physiques imposées par la structure de l'intasome, mais également grâce à des facteurs de ciblage cellulaires qui vont interagir avec à la fois l'intasome et des composants du nucléosome.

Dans ce projet de thèse, nous avons pu mettre en évidence une nouvelle interaction hôte-pathogène entre l'IN du VIH-1 et la queue d'histone H4 (une des protéines constituant le nucléosome). Ce projet s'est ainsi focalisé autour de cette interaction et a permis de :

- Démontrer l'importance de l'interaction entre l'IN du VIH-1 et la queue d'histone H4 lors du cycle viral et plus précisément pour l'étape d'intégration, validant ainsi cette interaction comme une nouvelle interaction hôte-pathogène.
- D'identifier que la queue d'histone H4 est un partenaire essentiel de l'intasome du VIH-1 pour qu'il puisse s'ancrer sur le nucléosome.
- Développer une nouvelle stratégie antivirale visant à bloquer cette interaction dans les cellules infectées grâce à des composés chimiques.

Mots clés : Rétrovirus, VIH-1, intégrase, intégration, chromatine, nucléosome, sélectivité

Functional study of the HIV-1 intasome - nucleosome interaction : the H4 histone tail as a new partner of integration

Abstract: HIV-1 integrase (IN) catalyzes the insertion of the viral genome into the host cell chromatin. This step is crucial for the virus for its efficient replication, integration is thus of interest to target for antiviral strategies. Understanding the mechanisms involved in integration is important in order to develop efficient tools to fight the virus.

Retroviral integration is catalyzed by the intasome, an oligomer of IN and viral DNA. Intasomes integrate onto nucleosomes, composed of DNA wrapped around histone proteins, over naked DNA.

In this thesis project, we have identified a new host-pathogen interaction between HIV-1 IN and the H4 histone tail. The topic of the project was then focus on this interaction and has highlighted:

- The importance of the HIV-1 IN – H4 histone tail interaction for the viral cycle, especially onto the integration step, validating a new host-pathogen interaction.
- The identification of the H4 histone tail as an essential partner for HIV-1 intasome for its anchoring onto nucleosomes.
- The development of a novel antiviral strategy aiming to block this interaction in infected cells using chemical compounds

Keywords: Retrovirus, HIV-1, integrase, integration, chromatin, nucleosome, selectivity

SOMMAIRE

SOMMAIRE.....	1
ABREVIATIONS.....	13
INTRODUCTION	23
1 Infection par le VIH-1	25
1.1 Le syndrome d'immunodéficience acquise	25
a) Historique.....	25
b) Epidémiologie	27
c) Physiopathologie.....	27
d) Diagnostic.....	29
1.2 Le virus de l'immunodéficience humaine 1.....	29
a) Classification et variabilité	29
b) Structure du virus.....	31
c) Génome du virus	31
d) Les protéines virales.....	33
e) Cycle de réplication.....	41
1.3 Stratégies antirétrovirales	47
a) Prévention.....	49
b) Traitements, résistance et réservoir	49
c) Essais Vaccinaux.....	55
2 L'intégration du VIH-1	57
2.1 Structure de l'IN du VIH-1.....	57
a) Le domaine N-terminal (NTD).....	57
b) Le corps catalytique (CCD)	57
c) Le domaine C-terminal (CTD).....	57
d) Les intasomes rétroviraux.....	59
2.2 Mécanisme d'action de l'IN	61
a) Le 3' processing.....	61
b) Le transfert de brin	61
c) Réparation post-intégration	63
d) Les produits non intégrés.....	63
2.3 L'intégration <i>in vitro</i>	63
a) Test de processing et de transfert de brin.....	65

b)	Le test d'intégration concertée.....	65
c)	La désintégration	67
2.4	Activités non catalytiques de l'IN	67
a)	Rôle de l'IN dans la transcription inverse	67
b)	Rôle de l'IN dans l'import nucléaire.....	67
c)	Rôle de l'IN dans les phases tardives du cycle viral	69
2.5	Les cofacteurs de l'intégration	69
3	La chromatine	75
3.1	Généralités.....	75
a)	Composition	75
b)	Le nucléosome : substrat final de l'intégration	75
3.2	Les modifications post-traductionnelles des queues d'histones	77
a)	La méthylation	77
b)	L'acétylation.....	79
c)	La phosphorylation	79
d)	L'ubiquitination	79
3.3	Dynamique de la chromatine	79
4	La sélectivité d'intégration	83
4.1	L'import nucléaire.....	83
4.2	Choix du site final	83
4.3	Impact de la structure de l'ADN nucléosomal.....	87
4.4	Impact de la condensation de la chromatine	87
5	Objectifs de l'étude.....	91
MATERIELS ET METHODES		93
1	Matériels	95
1.1	Culture cellulaire et bactérienne	95
a)	Les cellules	95
b)	Les bactéries.....	95
1.2	Plasmides.....	97
a)	Plasmide d'expression de l'IN du VIH-1	97
b)	Plasmides pour la production des vecteurs lentiviraux.....	97
1.3	Oligonucléotides.....	99
1.4	Peptides	99

1.5	Anticorps.....	101
1.6	Nucléosomes	101
a)	Nucléosomes commerciaux.....	101
b)	Nucléosomes « tail less ».....	101
1.7	Composés chimiques	103
1.8	Virus.....	103
a)	Vecteurs lentiviraux pseudotypés VSVG.....	103
b)	Le VIH-1 sauvage.....	103
2	Méthodes	105
2.1	Culture bactérienne.....	105
a)	Culture.....	105
b)	Transformation	105
2.2	Culture cellulaire.....	105
a)	Culture.....	105
b)	Tests de toxicité	107
c)	Infection par les vecteurs lentiviraux.....	107
d)	Infection par le virus sauvage	109
2.3	Manipulation des acides nucléiques	109
a)	Purification de l'ADN de cellules infectées	109
b)	qPCR	111
c)	Séquençage des sites d'intégration	111
d)	Mutagenèse dirigée	111
2.4	Manipulation des protéines	111
a)	Production de l'IN du VIH-1	111
b)	Purification des intasomes.....	113
c)	Western Blot	113
2.5	Tests in vitro	115
a)	Far Dot Blot	115
b)	Thermophorèse.....	115
c)	Test d'intégration concertée.....	115
d)	Pull Down	117
e)	Test de 3' processing et de transfert de brins	117
f)	Test d'intégration avec les intasomes	117

g)	Pontages chimiques au BS3	119
2.6	Analyse <i>in silico</i>	119
a)	Modélisation de l'interaction entre le CTD de l'IN et le peptide H4K20me1	119
b)	Criblage virtuel	119
RESULTATS.....		121
Partie 1 : L'interaction entre l'IN du VIH-1 et la queue d'histone H4 est fonctionnelle		123
1	Contexte scientifique et résultats préliminaires	123
2	Identification de la zone d'interaction de l'IN avec la queue d'histone H4	125
3	Analyse structurale et modélisation moléculaire <i>in silico</i> de l'interaction entre le CTD de l'IN et la queue d'histone H4	127
4	Recherche des résidus du CTD impliqués dans l'interaction avec la queue d'histone H4	129
5	Impact des mutations du CTD affectant l'interaction avec l'histone H4 sur les phases précoces de la réplication	131
6	Impact des mutations du CTD affectant l'interaction avec la queue d'histone H4 sur la sélectivité d'intégration	133
7	Conclusion et discussion.....	135
Partie 2 : Caractérisation biochimique de l'interaction IN-H4 au cours de l'intégration....		157
1	Contexte scientifique et objectifs	157
2	Effet des queues d'histones sur l'ancrage et l'intégration du VIH sur MN	157
3	Impact de l'interaction IN-queue d'histone H4 sur l'intégration <i>in vitro</i>	159
4	Etude <i>in vitro</i> du mécanisme d'action du peptide histone H4.....	163
4.1	Effet du peptide histone H4 sur le processing et le transfert de brin	163
4.2	Effet du peptide histone H4 sur des intasomes	165
4.3	Effet du peptide histone H4 sur la capture de l'ADN	167
4.4	Effet du peptide histone H4 sur l'oligomérisation de l'IN	167
5	Etude Structurale de l'effet du peptide H4 sur l'IN	169
6	Conclusion et discussion.....	171
Partie 3 : L'interaction IN-H4 comme cible antivirale		175
1	Contexte scientifique et objectifs	175
2	Etude du peptide H4K20me1 comme potentiel compétiteur de l'interaction entre l'IN et la queue d'histone H4	177
3	Effet des peptides histones sur les phases précoces de la réplication virale	177
4	Recherche de composés inhibant l'interaction entre l'IN et la queue d'histone H4 ...	179

4.1	Le criblage virtuel	181
4.2	Le criblage par alphascreen	181
5	Description des molécules identifiées en alphascreen	185
6	Caractérisation biochimique des molécules inhibant l'interaction IN-queue d'histone H4	187
7	Effet des molécules inhibant l'interaction IN-queue d'histone H4 sur la réplication virale	189
8	Conclusion et discussion.....	191
	Partie 4 : Conclusion générale et perspectives	193
1	L'interaction IN-queue d'histone H4 comme nouvelle interaction hôte-pathogène ..	193
2	L'activation de l'intasome au contact de la chromatine : une nouvelle étape de l'intégration ?	199
3	L'interaction IN-H4 comme nouvelle cible antivirale	205
4	Conclusion générale	205
	BIBLIOGRAPHIE.....	209

ABBREVIATIONS

A

ADN	Acide Désoxyribonucléique
ADNv	Acide Désoxyribonucléique viral
ALLINI	Allosteric intégrase inhibtor
ALV	Avian Leukosis Virus
APOBEC3G	Apolipoprotein B mRNA-Editing enzyme Catalytic polypeptide-like 3G
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messenger
ARNt	Acide ribonucléique de transfert
AS signal	Alphascreen signal
ATP	Adénosine triphosphate
AU	Arbitrary Unit
AZT	Azidothymidine

B

BAF	Barrier to Autointegration Factor
BET	Bromo and extra terminal domain
BRD	Bromodomain containing protein
BS3	BisSulfosuccinimidyl suberate
BSA	Bovin Serum Albumin

C

CA	Capside
CCD	Catalytic Core Domain
CCR5	C-C chemokine receptor type 5
CD4	Cluster of differentiation 4
CIC	Conserved Intasome Core
CMH	Complexe majeur d'histocompatibilité
CMV	Cytomegalovirus
CPI	Complexe de pré-intégration
CPSF6	Cleavage and polyadenylation factor 6
CRF	Circulating Recombinant Form
Cryo-EM	Cryo-electron microscopy
CSC	Cleaved Synaptic Complex
CTD	C-terminal domain
CXCR4	CXC chemokine receptor 4
CYPA	Cyclophilin A

D

DMEM	Dulbecco's modified minimal essential meduim
Do	Densité optique
DTG	Dolutegravir

E

eGFP	Enhanced Green Fluorescent Protein
ELISA	Enzyme-linked immunosorbent assay
Env	Enveloppe
ESCRT	Endosomal Sorting Complex Required for Transpor
EVG	Elvitegravir

F

FACT	Facilitates chromatin transcription
FDA	Food and Drug Administration
FEN-1	Flap endonuclease 1
FS	Full site
FSI	Full site integration

G

gag	Group-specific antigen
GCN2	General Control Nonderepressible 2
gp	Glycoprotéine
GST	Glutathione S-transferase

H

HEK	Human Embryonic Kidney
HeLa	Henrietta lacks
HIS	Histidine
HMGA1	High mobility group chromosomal protein A1
hMSL3	Human male-specific lethal-3 protein
HRP	Hepatoma-derived growth factor Related Protein
HSI	Half site integration

I

IBD	Integrase binding domain
IC	Intégration concertée
IC50	Half maximal inhibitory concentration
IN	Intégrase
INI1	Integrase interactor 1
INSTI	Integrase Strand Transfer Inhibitor

K

Kd	Constante de dissociation
KO	Knock Out

L

LAV	Lymphadenopathy Associated Virus
LB	Luria-Bertani
LEDGF/p75	Lens epithelium-derived growth factor
LEDGIN	LEDGF-IN interaction site
LT	Lymphocyte T
LTR	Long Terminal Repeat

M

MA	Matrice
MBT	Malignant brain tumor
MLV	Murine Leukemia Virus
MMTV	Mouse mammary tumor virus
MN	Mononucléosome
MOI	Multiplicity of infection
MP	Membrane plasmique
MPT	Modification post-traductionnelle

MVV	Maedi Visna Virus
N	
NC	Nucléocapside
NED	NTD extension domain
Nef	Negative factor
NHEJ	Non-Homologous End-Joining
NLS	Nuclear Localization Signal
NNRTI	Non Nucleoside Reverse Transcriptase Inhibitor
NPC	Nuclear Pore Complex
NRTI	Nucleoside Reverse Transcriptase Inhibitor
NTD	N-terminal Domain
NUP	Nucleoporin
O	
o.c.	Open circular
ODN	Oligodésoxyribonucléotide
P	
pb	Paire de base
PBMC	Peripheral blood mononuclear cell
PBS	Primer Binding Site
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PFV	Prototype Foamy Virus
PHA	Phytohémagglutinine
PPT	Poly Purin Tract
PR	Protéase
PrEP	Prophylaxie pré-exposition
Q	
qPCR	Quantitative PCR
R	
RAL	Raltégravir
RAV	Rous Associated Virus
rev	Regulator of expression of virion proteins
RMN	Résonance Magnétique Nucléaire
RRE	Rev Response Element
RSV	Rous Sarcoma Virus
RT	Reverse Transcriptase
RTC	Reverse Transcriptase Complex
S	
s.c.	Super coiled
SH3	Sarc homology 3 domain
SIDA	Syndrome d'Immunodéficience Acquise
siRNA	Small interfering RNA
SIV	Simian Immunodeficiency Virus
SSC	Stable Synaptic Complex
STC	Strand Transfer Complex

SVF	Serum de Veau Foetal
SWI/SNF	Switch/Sucrose Non Fermenting
T	
TALEN	Transcription activator-like effector nuclease
TAR	Trans-Activating Response Element
Tat	Transactivateur de la transcription
TCC	Target Capture Complex
TNPO3	Transportine 3
TL	Tail Less
V	
Vif	Viral infectivity factor
VIH	Virus de l'immunodéficience Humaine
Vpr	Viral protein R
Vpu	Viral protein U
VSV	Virus de la Stomatite Vésiculaire
W	
WDSV	Walleye Dermal Sarcoma Virus
WT	Wild Type

INTRODUCTION

INTRODUCTION

1 Infection par le VIH-1

1.1 Le syndrome d'immunodéficience acquise

a) Historique

Le virus de l'immunodéficience humaine (VIH), est un rétrovirus pathogène de l'homme responsable du syndrome d'immunodéficience acquise (SIDA) chez les personnes qu'il infecte. Cette maladie virale va aboutir à la destruction des cellules immunitaires jusqu'au stade où l'individu infecté ne pourra plus se défendre contre les maladies opportunistes s'il n'est pas pris en charge.

Les premiers signes annonciateurs de cette épidémie ont été observés pour la première fois en 1981, où le Center for Disease Control d'Atlanta a rapporté quelques cas de graves pneumonies dues à des infections à *Pneumocystis carinii* et des sarcomes de Kaposi. Dans les semaines qui suivent, plusieurs cas proches commencent à intriguer chercheurs et médecins, seulement, très peu de choses sont connues à propos de cette maladie. Seuls points communs entre ces patients : le développement d'une grave immunodéficience qui semble se transmettre par voie sexuelle et sanguine, touchant principalement les homosexuels, des personnes transfusées et des utilisateurs de drogues injectables.

Ces caractéristiques laissent les scientifiques penser que cette pathologie est causée par un agent infectieux, et, dès 1982, plusieurs équipes de chercheurs à travers la planète se mobilisent afin d'identifier cet agent. Cette année-là, les symptômes de cette maladie sortent des frontières Américaines et sont observés, entre autres, en France chez des hémophiles transfusés. Le terme de SIDA sera alors utilisé pour la première fois pour décrire cette pathologie.

En 1983, l'analyse d'un échantillon de ganglions lymphatiques d'un patient par une équipe Française, a mis en évidence une activité de transcription inverse démontrant la présence d'un rétrovirus. Ce virus sera dans un premier temps appelé lymphadenopathy-associated virus¹ (LAV). Durant la même période, une équipe Américaine déclare avoir trouvé l'agent causal du SIDA et baptise le virus HTLV-III².

En 1985, le génome du virus est séquencé et prendra finalement le nom de VIH l'année suivante. Cette même année, un deuxième type de VIH est découvert : le VIH-2, moins pathogène que le VIH-1.

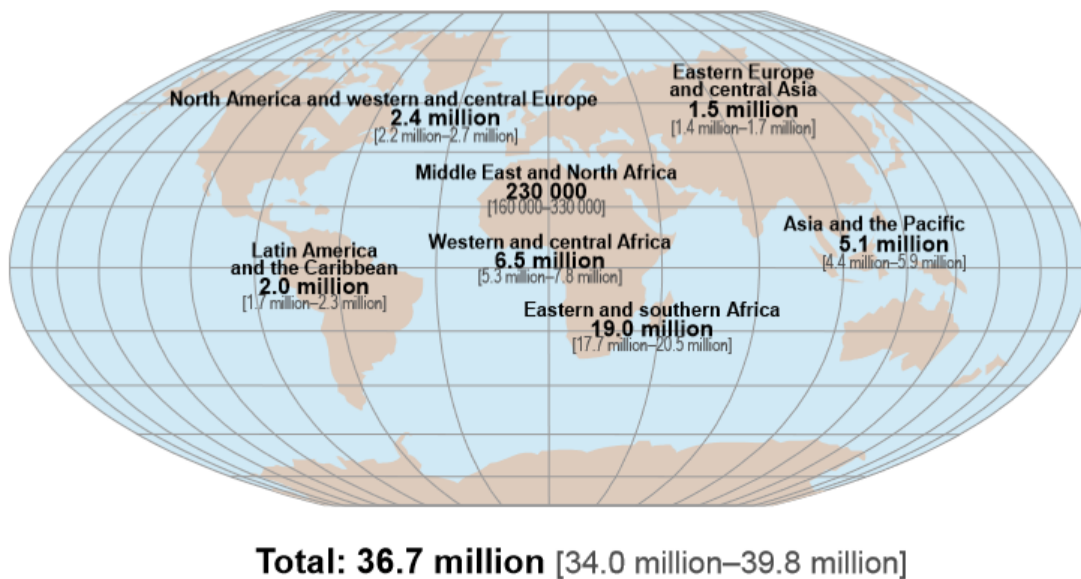


Figure 1 : Répartition géographique des personnes vivant avec le SIDA en 2016. D'après le rapport ONUSIDA de 2016, environ 36,7 millions de personnes sont infectées dans le monde.

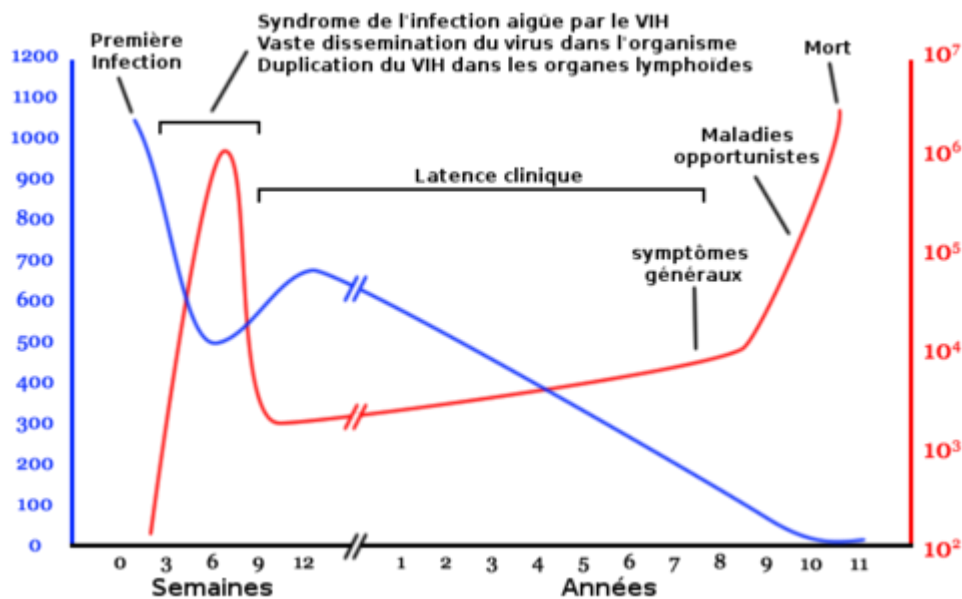


Figure 2 : Evolution de l'infection par le VIH chez une personne non traitée. En bleu est indiqué le nombre de cellules CD4+ par mm³ de sang et en rouge le nombre de copies virales par mL de plasma. Sont également indiquées les trois phases retrouvées classiquement chez un sujet infecté : la primo-infection, la phase asymptomatique et le stade SIDA. (<https://www.techno-science.net/illustration/Autres/Medecine/Graphe-infection-VIH.jpg>)

b) Epidémiologie

D'après les derniers rapports ONUSIDA de 2018, environ 36,7 millions de personnes étaient séropositives en 2016 avec 1,8 millions de nouvelles contaminations cette année-là et un million de morts dus à cette maladie. Depuis que le SIDA est considéré comme une pandémie (en 2002), il est estimé que plus de 25 millions de personnes sont décédées suite à une infection par le VIH. La répartition géographique des personnes infectées par le virus est très inégale à travers le globe (**Figure 1**). La région la plus touchée est l'Afrique subsaharienne, qui regroupe à elle seule 69% des personnes séropositives à travers le monde.

Cependant, la mise en place de campagnes de prévention et le développement de thérapies antirétrovirales ont très largement contribué à faire chuter de manière significative le taux de mortalité dû au SIDA, notamment depuis l'apparition des premières trithérapies en 1996. Malheureusement ces traitements ne sont pas accessibles pour tous, ce qui explique en partie pourquoi le SIDA est toujours aussi prédominant dans certaines régions du globe.

c) Physiopathologie

Le VIH infecte les cellules présentant des récepteurs CD4 à leur surface. Ainsi, les lymphocytes T CD4+ (LT CD4), les macrophages, les cellules dendritiques et les cellules microgliales cérébrales peuvent être infectées par le VIH. L'infection se déroule en plusieurs phases se succédant dans le temps (**Figure 2**) et aboutissant à la destruction progressive des cellules du système immunitaire. En absence de traitement on distingue classiquement trois phases au cours de l'infection :

- **La primo-infection**

C'est la première phase survenant suite à la contamination par le virus. Pendant cette période le virus se réplique de manière efficace et entraîne une destruction rapide des cellules CD4 qu'il infecte. Ceci provoque des symptômes variés chez les sujets infectés dans les semaines suivant la contamination, sous forme de syndrome pseudo grippal accompagné le plus souvent de fièvre. Durant cette phase, la réponse du système immunitaire parvient à contrôler la réplication virale et est suffisante pour maintenir un renouvellement à peu près stable des LT CD4.

- **La phase asymptomatique**

Cette phase peut durer de quelques mois à plusieurs années. Le virus continue de se multiplier massivement et la destruction progressive des cellules CD4 se poursuit jusqu'à atteindre un stade où le patient ne peut plus assurer une production suffisante de cellules immunitaires pour compenser celles détruites par cet état d'hyper activation immunitaire ou par le virus directement. Vient alors la phase SIDA.

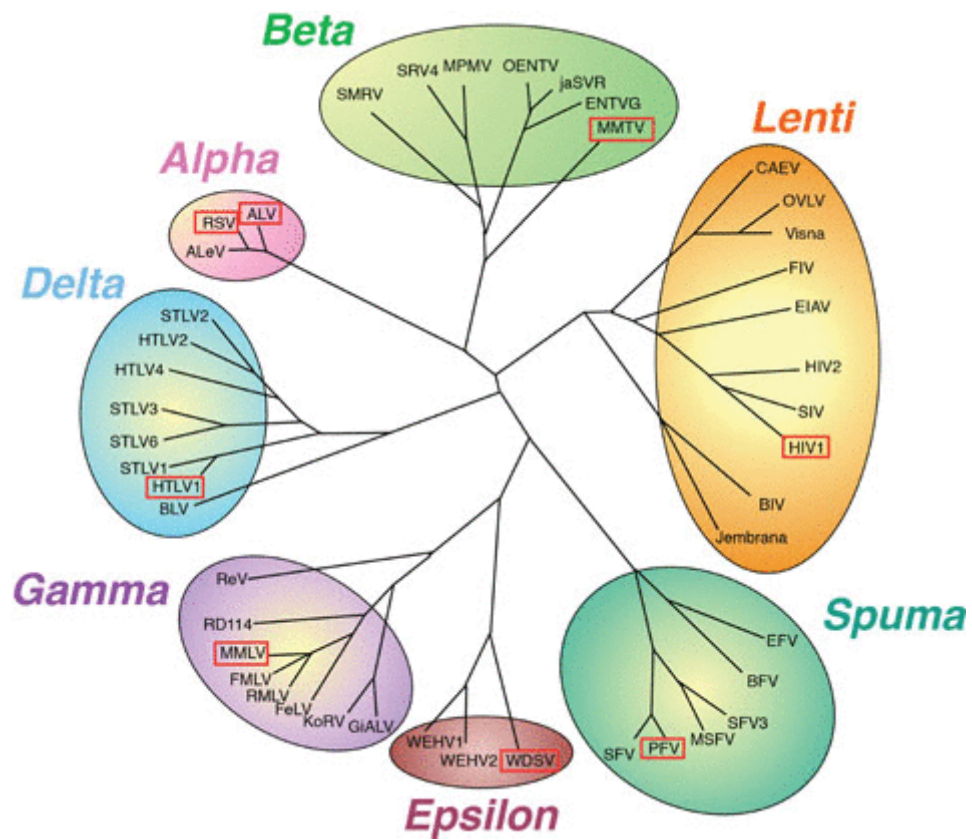


Figure 3 : Arbre phylogénétique des rétrovirus basé sur les alignements de séquences de la transcriptase inverse et de l'intégrase³.

- **La phase symptomatique ou SIDA**

Une virémie élevée et l'atteinte d'un seuil critique de LT CD4 (<200 par mm³ de sang) entraînent l'apparition des premiers signes cliniques de l'infection. Le sujet immunodéprimé ne peut plus contrôler les pathologies infectieuses et les maladies opportunistes qui conduisent au décès du patient qui n'a plus de moyens de défense contre ces attaques.

d) Diagnostic

La détermination du statut sérologique (positif ou négatif) au VIH d'une personne se fait en deux étapes. Dans un premier temps, la présence d'anticorps anti-VIH est recherchée par ELISA dans le sérum ou le plasma. Si ce résultat est négatif, l'individu est alors considéré comme séronégatif (cependant si l'infection est trop récente, le test peut ne pas détecter les anticorps). Si le résultat est positif, il doit dans un deuxième temps être confirmé par western blot que les anticorps sont bien spécifiques à une infection par le VIH afin d'éliminer d'éventuels faux positifs. Dans certains cas, la recherche de l'antigène p24 ou de l'ARN viral peut également s'avérer nécessaire.

1.2 Le virus de l'immunodéficience humaine 1

a) Classification et variabilité

Il existe des rétrovirus endogènes et exogènes. Les endogènes désignent une séquence du génome d'un organisme stable (qui se transmet verticalement) et ayant des analogies avec certains rétrovirus. Il est estimé que les rétrovirus endogènes humains représentent environ 8% du génome.

Le VIH-1 appartient à la famille des *Retroviridae*. Ce sont des virus enveloppés à ARN monocaténaire de polarité positive. On distingue 7 genres de rétrovirus exogènes au sein de cette famille (**Figure 3**) classés suivant divers critères tels que la morphologie, la pathogénicité et l'organisation du génome :

- Les Alpharétrovirus (ex : Avian Leukosis Virus, ALV)
- Les Betarétrovirus (ex : Mouse Mammary Tumor Virus, MMTV)
- Les Gammarétrovirus (ex : Murine Leukemia Virus, MLV)
- Les Deltarétrovirus (ex : Human T-Lymphotropic Virus, HTLV)
- Les Epsilonrétrovirus (ex : Walleye Dermal Sarcoma Virus, WDSV)
- Les Lentivirus (ex : Human Immunodeficiency Virus, HIV)
- Les Spumavirus (ex : Prototype Foamy Virus, PFV)

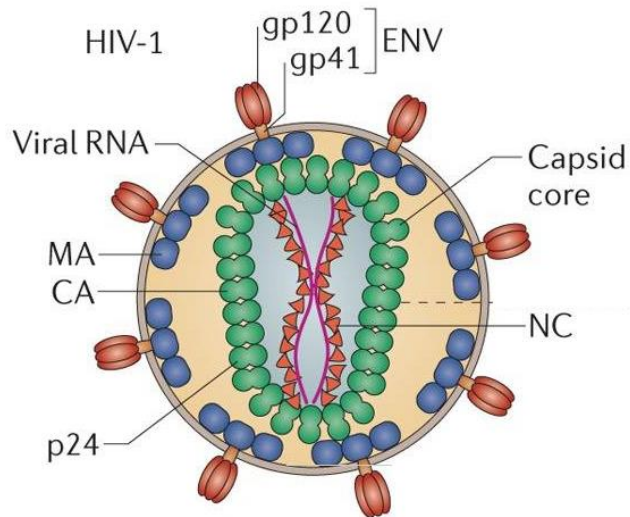


Figure 4 : Représentation schématique d'une particule virale du VIH mature. MA : matrice ; CA : capsid ; NC : nucléocapside⁴.

Des infections zoonotiques ont permis l'émergence de plusieurs groupes différents de virus au sein du VIH-1. Parmi ces groupes on distingue : le groupe M (Major), le groupe O (Outlier), le groupe N (Non M, Non O) et le groupe P.

Le groupe M est responsable de la pandémie de SIDA et comporte 9 sous-types (A, B, C, D, F, G, H, J et K) ainsi que plus de 80 formes recombinantes entre ces différents sous-types que l'on appelle CRF (Circulating Recombinant Form) résultant de recombinaisons dans le génome viral de plusieurs sous-types de VIH-1 lors de co-infections chez un même individu⁵.

b) Structure du virus

La particule virale (**Figure 4**) est de forme icosaédrique avec un diamètre d'environ 100 nm. Son enveloppe est composée d'une bicouche lipidique provenant de la cellule productrice ainsi que des protéines virales d'enveloppe : les glycoprotéines de surface gp120 et gp41 permettent aux virions de reconnaître le récepteur CD4 ainsi que les corécepteurs CCR5 ou CXCR4 des cellules cibles du VIH. Sur la face interne de l'enveloppe se trouve la matrice (MA, p17) formant un réseau stabilisant la particule virale. Vient ensuite la capside (CA, p24), de forme conique, elle s'organise en hexamères qui permettent la protection de l'ARN. Au sein de cette capside on retrouve également la nucléocapside (NC, p7) qui interagit directement avec l'ARN viral le protégeant ainsi des nucléases cellulaires. Enfin, diverses protéines virales se trouvent aussi dans la capside telles que la transcriptase inverse (RT, p66/p51), l'intégrase (IN, p32), la protéase (PR, p10) ainsi que des protéines dites « régulatrices ».

c) Génome du virus

Le génome du VIH-1 est composé de deux copies d'ARN simple brin de polarité positive d'environ 9200 nucléotides. Après transcription inverse de cet ARN par la RT, l'ADN double brin nouvellement synthétisé sera par la suite intégré dans le génome de la cellule, on parle alors de provirus. Les extrémités de ce provirus sont flanquées par les séquences LTR (Long Terminal Repeat) faisant environ 600 paires de base (pb). Ces LTR sont composés des régions U3, R et U5 qui jouent un rôle central dans l'intégration et l'expression du provirus.

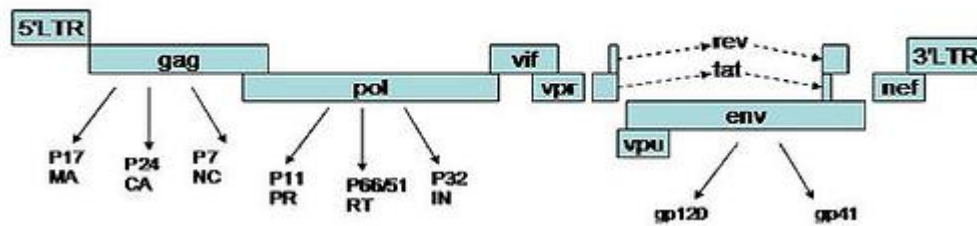


Figure 5 : Représentation de l'organisation génomique du VIH-1. Le génome du VIH-1 est flanqué à ses extrémités des séquences LTR en 5' et en 3'. Le VIH-1 code pour 9 gènes au total qui permettent la production des différentes protéines virales de structure, les enzymes et protéines régulatrices et accessoires⁶.

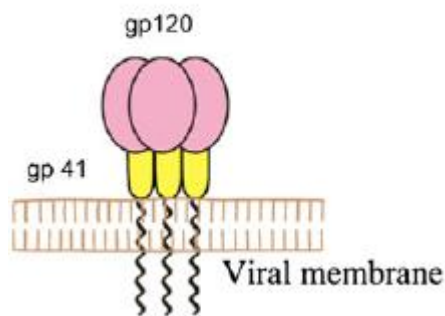


Figure 6 : Représentation schématique de l'organisation des protéines virales gp120 et gp41. La protéine gp41 (en jaune) est transmembranaire alors que la protéine gp120 (en rose) n'est pas en contact avec la membrane virale. Ces deux protéines s'agencent en trimères de gp41 et de gp120 pour former des complexes appelés « spikes »⁷.

L'ADN proviral (**Figure 5**) code pour neuf gènes qui peuvent être répartis en trois catégories : les gènes codant pour les protéines structurales et non structurales (*gag*, *pol* et *env*), les gènes codant pour les protéines régulatrices (*tat* et *rev*) et pour finir les gènes codant pour les protéines accessoires (*vpu*, *vpr*, *vif* et *nef*)⁸.

- **Le gène *gag* (group-specific antigen)**

Ce gène code pour le précurseur polyprotéique gag (Pr55). Cette polyprotéine sera maturée dans les virions par clivage par la protéase virale, en protéine de matrice, de capside, de nucléocapside ainsi que des peptides SP1, SP2 et p6.

- **Le gène *pol***

Ce gène code pour le précurseur gag-pol (Pr160), qui après maturation par la protéase donnera les enzymes intégrase, transcriptase inverse et protéase.

- **Le gène *env* (enveloppe)**

Ce gène code pour la polyprotéine gp160 qui sera clivée par une protéase cellulaire en deux protéines : la glycoprotéine de surface gp120 et la glycoprotéine membranaire gp41.

Les protéines régulatrices et accessoires portent le même nom que le gène dont elles sont issues et sont détaillées dans la section suivante.

d) Les protéines virales

- **Les glycoprotéines d'enveloppe**

Les glycoprotéines d'enveloppe du VIH sont au nombre de deux : la gp120 et la gp41. Elles sont impliquées dans la reconnaissance des cellules cibles du virus grâce à des interactions spécifiques entre ces glycoprotéines et les récepteurs cellulaires CD4 et corécepteurs CXCR4 et CCR5⁹. Suite à ces interactions, des changements conformationnels des glycoprotéines vont permettre la fusion des membranes virale et cellulaire aboutissant à l'entrée du virus dans sa cellule cible. Ces deux glycoprotéines s'associent à la surface du virus pour former des complexes composés de trimères de gp120 et de gp41 (**Figure 6**).

La gp120 est une protéine avec des régions constantes (C1 à C5) et des régions variables (V1 à V5). Ces régions variables permettent l'interaction avec les corécepteurs cellulaires et sont en grande partie responsables de l'échappement au système immunitaire grâce à leur variabilité et leur forte glycosylation.

La gp41 est une protéine transmembranaire qui contient trois domaines majeurs : le domaine externe, transmembranaire et interne¹⁰. Le domaine externe contient une portion appelée « peptide de fusion », qui est exposée une fois que la gp120 ait subi des changements conformationnels suite à l'interaction avec le CD4 de la cellule cible. Une fois exposé, le peptide de fusion va s'ancrer dans la membrane cellulaire débutant ainsi le processus de fusion avec la membrane virale.

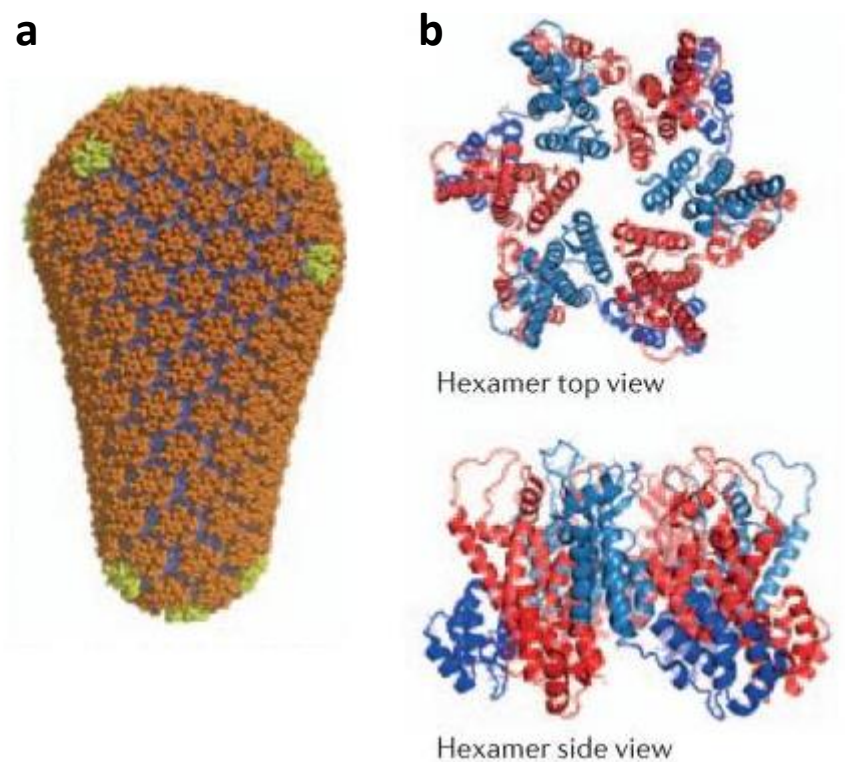


Figure 7 : Organisation et agencement des oligomères de capsid. **(a)** Représentation du core conique que forme la capsid, composé d'hexamères de capsid (en orange) et de pentamères (en jaune). **(b)** Vues du dessous et de côté d'un hexamère de capsid (en bleu : le domaine C-terminal ; et en rouge : le domaine C-terminal)⁴.

- **Les protéines structurales**

La matrice (MA)

Dans les virions matures la protéine de matrice interagit et recouvre la surface interne de l'enveloppe virale. Elle possède trois séquences de localisation : une d'import nucléaire, une d'export nucléaire et une de transport à la membrane^{11,12}. Cette dernière joue un rôle important dans les étapes tardives de la réplication virale, en dirigeant l'ARN du virus au site d'assemblage du virion au niveau de la membrane plasmique cellulaire.

La capside (CA)

Lors de la maturation du précurseur gag par la protéase virale, les monomères de CA ainsi générés vont s'assembler en hexamères et en pentamères pour former le core viral¹³, structure close protégeant le génome du virus, facilement reconnaissable grâce à sa forme conique⁴ (**Figure 7**). Il est supposé que la CA sous forme de core joue deux rôles principaux : premièrement en assurant le bon déroulement de la transcription inverse en maintenant la RT et l'ARN viral proches dans un environnement clos. L'étape de transcription inverse requiert de nombreux sauts de brins par la RT entre l'ARN et l'ADN nouvellement synthétisé, et il est supposé que ceci se fait optimalement si tous les composants nécessaires à cette étape sont tous facilement à portée les uns des autres. Deuxièmement, le core viral sert de protection au génome du virus, l'empêchant d'être détecté par des facteurs cellulaires initiateurs de réponse antivirale¹⁴.

La manière dont la décapsidation se déroule n'est pas encore précisément déterminée, en revanche il a été démontré qu'il reste de la CA au sein du complexe de pré-intégration (CPI) dans le noyau, et plusieurs rôles de cette CA ont été identifiés au cours du cycle viral, notamment lors de l'import nucléaire et de l'intégration, révélant un rôle bien plus important de la CA que seulement l'encapsidation du génome viral⁴.

Une des caractéristiques des lentivirus est de pouvoir infecter des cellules qui ne sont pas en division, car leur CPI est activement transporté dans le noyau de la cellule par les pores nucléaires. Récemment, il a été démontré que la CA jouait un rôle déterminant lors de l'import nucléaire, en interagissant notamment avec une protéine cellulaire : CPSF6¹⁵ (Cleavage and polyadenylation factor 6). Cette interaction est également déterminante pour le choix du site d'intégration¹⁶, et sera donc détaillée plus en profondeur par la suite. La CA interagit également avec d'autres protéines cellulaires impliquées dans l'import nucléaire comme TNPO3¹⁷, NUP153¹⁸ et Nup358¹⁹. Etant donné le rôle prédominant de la CA lors du cycle de réplication du virus, des stratégies antivirales commencent à voir le jour en ciblant cette protéine.

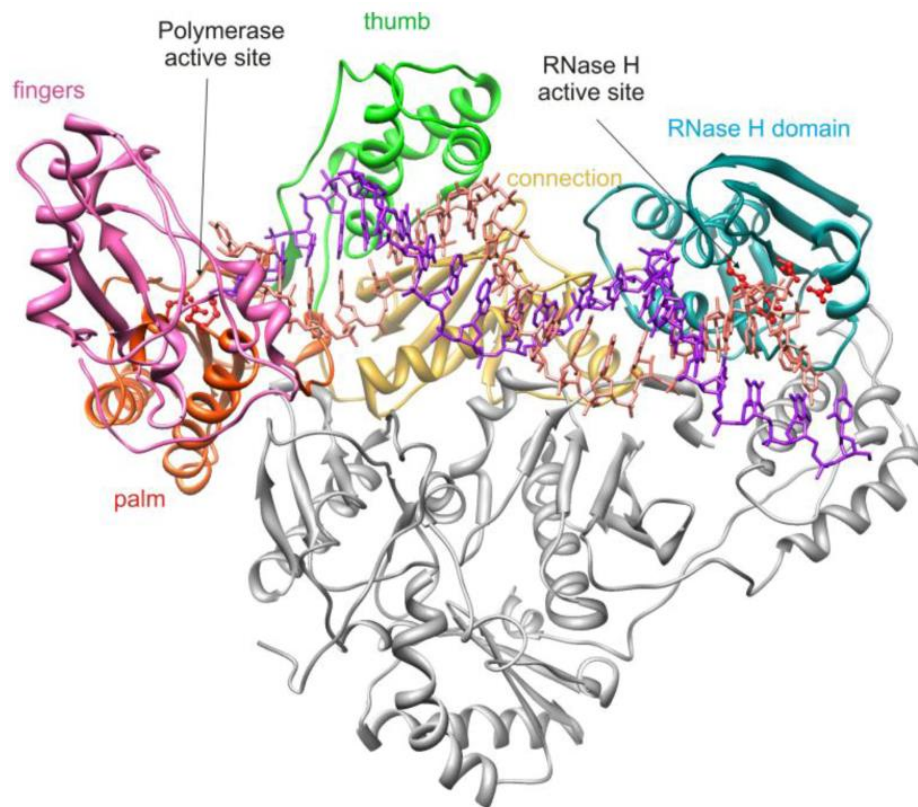


Figure 8 : Structure de la transcriptase inverse du VIH-1 en complexe avec un acide nucléique. La sous-unité p51 est schématisée en gris et les sous-domaines de la sous-unité sont représentés en couleurs²⁰.

La nucléocapside (NC)

La NC est une protéine chaperonne des acides nucléiques, ce qui signifie qu'elle peut influencer sur leur structure en leur faisant adopter des conformations plus stables. Cette protéine contient deux domaines à doigt de zinc à motif CCHC²¹ qui confèrent à la NC une forte affinité pour l'ARN du virus. Le rôle de la NC est multiple : elle protège, participe à l'assemblage, l'internalisation et la dimérisation de l'ARN viral²². Elle est également importante pour la transcription inverse en assurant la spécificité et l'efficacité de cette étape.

La protéine p6

La protéine p6 est codée par la région la plus variable du gène *gag*. Elle est impliquée dans l'incorporation de Vpr dans les virions et facilite également leur bourgeonnement lors de la formation des nouveaux virus²³.

- **Les protéines non structurales**

La protéase (PR)

La PR du VIH fait partie de la famille des aspartates protéases, qui est active sous forme d'homodimère. Son site actif se trouve à l'interface des dimères, il est constitué d'une triade catalytique Asp-Thr-Gly caractéristique des protéases de cette famille²⁴. Son rôle est de maturer les précurseurs polyprotéiques gag et gag-pol en les hydrolysant à des sites spécifiques, générant ainsi des protéines fonctionnelles pour former des virus matures et infectieux.

La PR virale peut également contribuer à la déplétion des cellules infectées en clivant des protéines cellulaires comme Bcl-2 et la caspase-8, générant ainsi des signaux déclencheurs de l'apoptose²⁵.

La transcriptase inverse (RT)

La RT est active sous forme d'hétérodimère, composée des sous-unités p51 et p66^{26,27} (**Figure 8**). La sous-unité p51 est issue du clivage du dimère p66/p66 par la PR dans le virion. Le rôle de la RT est de convertir l'ARN simple brin du virus en un ADN double brin, structure compatible à l'intégration dans le génome de l'hôte. Pour cela, la RT possède trois activités enzymatiques : une activité ADN polymérase ARN dépendante, une activité ADN polymérase ADN dépendante et une activité RNase H.

L'intégrase (IN)

Cette enzyme appartient à la famille des polynucléotidyl-transférases. L'IN du VIH se subdivise en trois domaines : le domaine N-terminal, le core catalytique et le domaine C-terminal. Même si l'IN est bien connue pour jouer un rôle central dans l'intégration rétrovirale, elle est également impliquée dans d'autres étapes du cycle du virus aussi bien dans les étapes précoces que tardives. L'intégrase étant au cœur de ce projet de thèse, elle sera décrite plus en profondeur dans la partie 2 : L'intégration rétrovirale.

- **Les protéines régulatrices**

Tat (trans-activator of transcription)

Tat est une protéine qui permet d'augmenter fortement le niveau de transcription du génome viral. Lors des premiers événements de transcription du provirus, Tat n'est pas présente, la transcription virale est faible. Mais une fois la protéine traduite, elle va venir se fixer sur une région de l'ARN du virus nouvellement transcrit : TAR^{28,29} (Trans-Activating Response Element). Ceci va permettre le recrutement de plusieurs facteurs de transcription, augmentant ainsi le niveau de production d'ARNm viral.

Rev (regulator of expression of virion proteins)

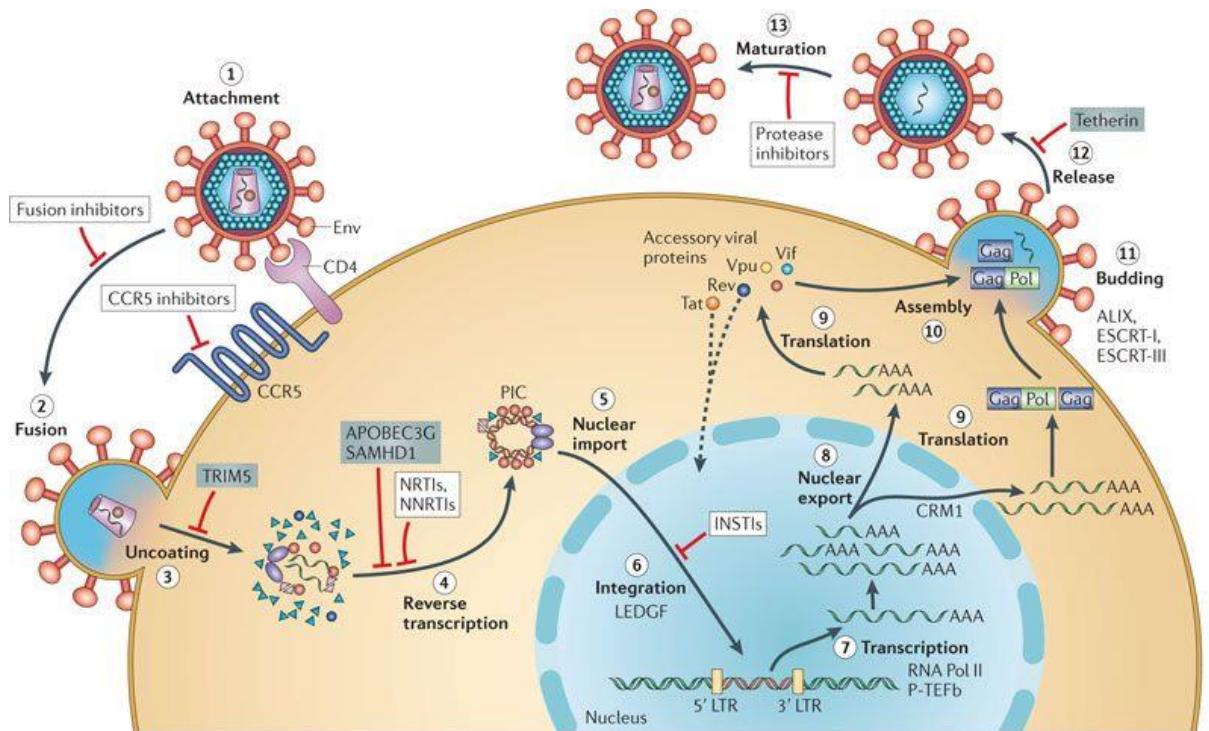
La protéine Rev est impliquée dans l'expression des protéines virales en favorisant l'export des ARN messagers (ARNm) viraux non ou mono-épissés. Rev contient à la fois des séquences de signal d'export et d'import nucléaire, ainsi elle est capable de se fixer spécifiquement aux ARNm dans le noyau au niveau de la séquence RRE (Rev Response Element) de ces derniers et de les exporter dans le cytoplasme où ils seront traduits pour produire les protéines virales³⁰.

- **Les protéines accessoires**

Vpu (viral protein U)

Vpu assure deux fonctions distinctes. La première est d'induire la dégradation du récepteur cellulaire CD4 des cellules infectées³¹. Pour cela, Vpu lie le CD4 à un complexe ubiquitine ligase, provoquant ainsi l'ubiquitination du CD4 et sa dégradation par le protéasome au niveau du réticulum endoplasmique. Ceci empêche l'interaction entre les protéines Env du virus et le CD4, car ces interactions peuvent interférer avec l'assemblage correct des protéines Env dans le virion. Cette fonction limite également la réinfection de la cellule.

La seconde fonction de Vpu est de favoriser la libération des virions des cellules infectées en neutralisant un facteur de restriction cellulaire : BST-2, qui en absence de Vpu, retient les virions sur la membrane plasmique de la cellule^{32,33}.



Nature Reviews | Microbiology

Figure 9 : Représentation schématique du cycle de réplication du VIH-1. Le cycle viral débute par l'entrée du virus dans la cellule cible par reconnaissance de récepteurs et corécepteurs cellulaires (1). Après fusion des membranes viral et cellulaire (2), la capside est partiellement désassemblée (3) et l'ARN du virus est converti en ADN lors de la transcription inverse (4). L'ADN viral va ensuite être importé dans le noyau (5) où il sera alors intégré dans le génome de la cellule (6). Après transcription et traduction des protéines virales par la machinerie cellulaire, de nouveaux virions vont s'assembler au niveau de la membrane plasmique où ils seront libérés par bourgeonnement (7 à 12). Le cycle viral s'achève par la maturation des précurseurs polyprotéiques (13) rendant les virions infectieux. Certaines cibles thérapeutiques antivirales ainsi que des protéines de l'hôte limitant l'infection sont également mentionnées³⁴.

Vpr (viral protein R)

Vpr a la capacité d'induire un arrêt du cycle cellulaire en phase G2³⁵. Cet arrêt entraîne une activation de la machinerie de réparation de l'ADN qui est supposé faciliter l'intégration de l'ADN viral, même si le rôle de Vpr dans ce processus n'est pas encore tout à fait clair.

Vpr est retrouvée dans les virions et est donc présente lors des phases précoces de la réplication virale. Il a été démontré que Vpr joue un rôle dans l'import nucléaire du CPI dans les macrophages³⁶ (cellules qui ne se divisent pas), permettant ainsi au virus de se répliquer efficacement dans ces cellules.

Vif (viral infectivity factor)

La protéine Vif participe à la dégradation d'un facteur de restriction à l'infection du VIH-1 : APOBEC3G, qui inhibe l'étape de transcription inverse³⁷. En induisant l'ubiquitination d'APOBEC3G et sa dégradation par le protéasome, Vif permet au virus d'effectuer la reverse transcription de manière efficace.

Nef (negative regulatory factor)

Les principales fonctions de Nef sont de perturber la transduction de signaux cellulaires et d'inhiber l'expression de plusieurs protéines de surface impliquées dans les défenses immunitaires (CD4, CMH-1, CMH-2...), ce qui a un impact drastique sur la virulence de l'infection³².

e) Cycle de réplication

Le cycle de réplication du VIH-1 (**Figure 9**) se divise en deux grandes phases : les phases précoces, allant de l'entrée du virus à l'intégration de l'ADN viral, et les phases tardives, allant de la transcription du provirus à la maturation des virions.

- **Les étapes précoces**

L'entrée du virus

La première étape du cycle de réplication du VIH-1 est son entrée dans la cellule cible. Dans un premier temps, l'interaction entre la glycoprotéine de surface gp120 du virus avec le récepteur cellulaire CD4³⁸ va entraîner un changement de conformation de la gp120, exposant un nouveau site d'interaction avec un autre récepteur cellulaire (CCR5 ou CXR4 suivant le tropisme du virus). Ceci déclenche l'insertion du peptide de fusion contenu dans la gp41 à l'intérieur de la membrane cellulaire puis à la fusion complète des deux membranes³⁹.

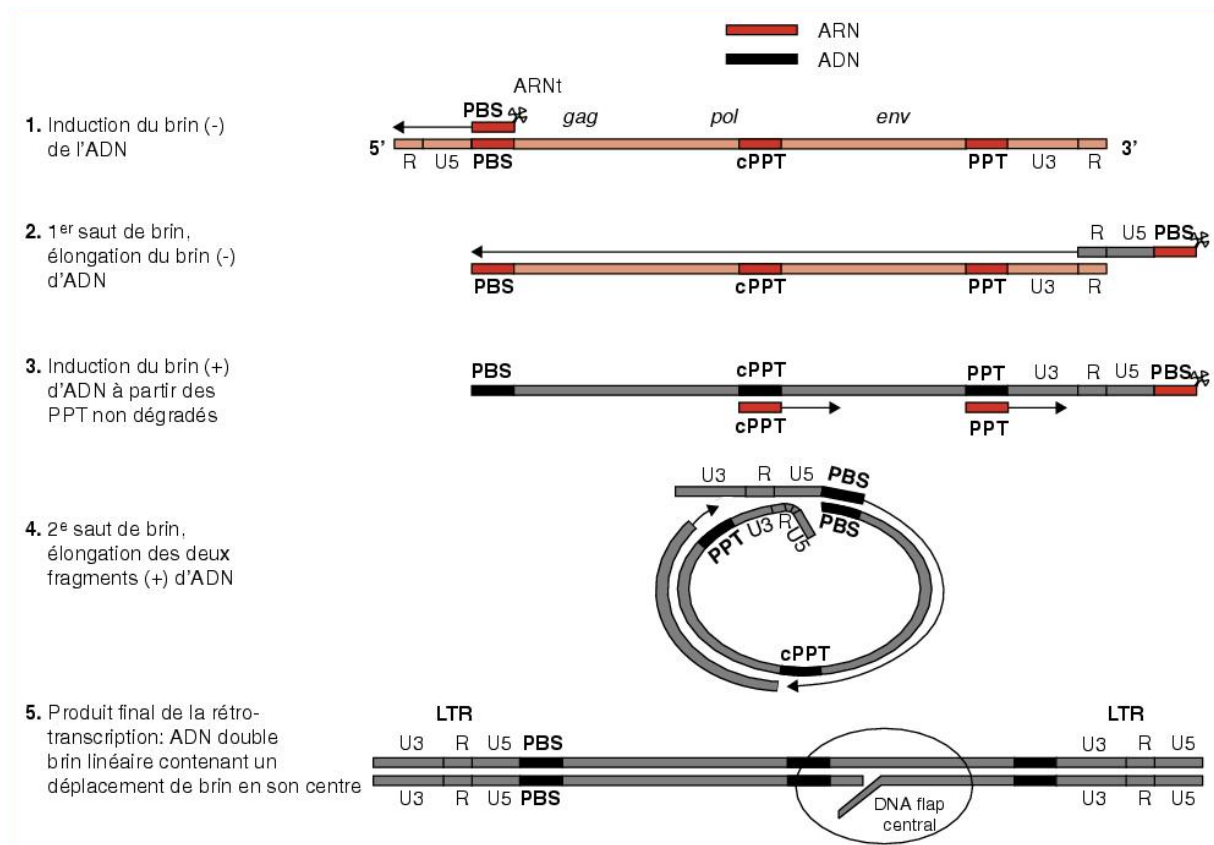


Figure 10 : Les différentes étapes de la transcription inverse du VIH-1⁴⁰.

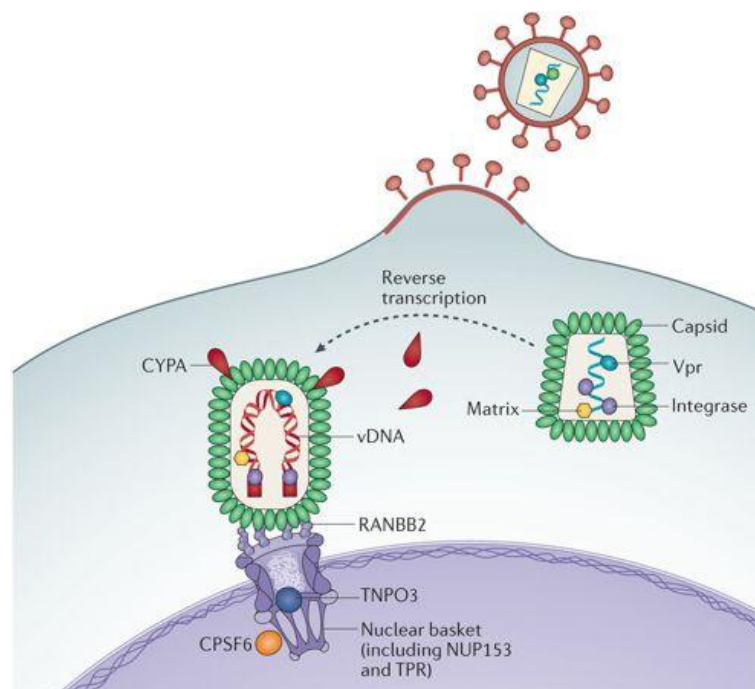


Figure 11 : Représentation schématique des protéines impliquées dans l'import nucléaire. La cyclophiline A (CYPA) et RANBP2 sont impliqués dans l'import nucléaire du complexe de pré-intégration (CPI) du côté du cytosol. La transportine 3 (TNPO3) interagit avec le CPI à l'intérieur du pore nucléaire. Nup153 et CPSF6 vont finaliser l'import nucléaire du CPI et vont également influencer sur le choix du site d'intégration⁴¹.

La transcription inverse

Le moment exact de l'initiation de la transcription inverse n'est pas encore clairement établi : certaines études ont montré qu'il y avait de l'ADN contenu dans les virions⁴²⁻⁴⁴, alors que d'autres pensent qu'elle débute une fois la capsid libérée dans le cytoplasme. La réaction est initiée par la fixation d'un ARNt cellulaire au niveau de la séquence PBS (primer binding site) située dans la région 5' de l'ARN viral (**Figure 10**). Cet ARNt va servir d'amorce pour la RT qui va dans un premier temps synthétiser un fragment du brin d'ADN négatif jusqu'à l'extrémité 5' de l'ARN. La portion ARN du duplex ARN/ADN va être dégradée par l'activité RNase H de la RT, et la séquence ADN va effectuer un premier saut de brin au niveau de l'extrémité 3' de l'ARN grâce à la présence des séquences R répétées sur les extrémités de l'ARN. L'élongation du brin d'ADN négatif va se poursuivre sur toute sa longueur, la RT dégradant à nouveau l'ARN sur les duplex ARN/ADN, mis à part les très stables régions PPT (poly purine tract) qui serviront ensuite d'amorce pour la synthèse du brin d'ADN positif. Après que l'ARNt ait été éliminé, un second saut de brin au niveau des séquences PBS des deux brins d'ADN va permettre la finalisation de la synthèse de l'ADN double brin linéaire avec les séquences LTR à ses deux extrémités⁴⁵.

A noter que toutes ces étapes sont réalisées au sein du complexe de transcription inverse (RTC), qui comprend diverses protéines virales pouvant influencer sur le bon déroulement de la transcription inverse. Une fois l'ADN double brin synthétisé, son association avec diverses protéines virales et cellulaires va former le complexe de pré-intégration⁴⁶.

Le mécanisme de désassemblage du core de capsid est mal connu, mais il est présumé qu'il se fait assez vite après l'entrée du virus, de manière partielle, car des traces de CA sont retrouvées dans le RTC. De plus, l'import nucléaire requiert qu'il y ait de la CA dans le CPI⁴.

L'import nucléaire

L'import nucléaire du CPI des lentivirus se fait de manière active par les pores nucléaires, ce qui leur permet d'infecter des cellules qui ne sont pas en division. Etant donné que la taille du CPI excède la taille des pores nucléaires, ce processus est énergie-dépendant, et est régulé par plusieurs protéines virales et cellulaires (**Figure 11**). L'environnement dans lequel va se faire cette étape est très important car cela va influencer sur le choix du site d'intégration⁴¹.

Un des déterminants viraux qui a le plus d'impact sur l'import nucléaire est la capsid. Des virus chimériques VIH-1 contenant la CA de MLV (qui est dépendant du cycle cellulaire) ont été montrés comme ayant un impact sévère sur l'import nucléaire et devenaient par conséquent dépendants du cycle cellulaire à leur tour⁴⁷. La CA interagit avec de nombreuses protéines cellulaires telles que la nucléoporine (NUP) 153^{18,48,49}, NUP358^{19,50}, CPSF6⁵¹⁻⁵³ et la cyclophiline A⁵⁴ (CYPA) dont chacune d'entre elles ont également été montrées comme influençant le choix du site d'intégration. La déplétion de NUP358 entraîne une diminution des sites d'intégration dans les régions denses de la chromatine, tout comme un mutant de

la CA (N74D) n'interagissant plus avec CPSF6⁵⁰. En revanche, la déplétion de CYPA induit un phénotype opposé : une augmentation des sites d'intégration dans les régions denses de la chromatine, même si le mécanisme de CYPA n'est pas encore tout à fait clair⁵⁵.

L'implication de CPSF6 dans l'import nucléaire et la sélectivité d'intégration sera discutée de manière plus approfondie dans la partie 2 : L'intégration rétrovirale, car cette protéine a récemment été identifiée comme étant cruciale dans ces étapes du cycle viral.

L'intégration

Une fois arrivé dans le noyau, le CPI va s'orienter vers un environnement favorable où s'ancrer à la chromatine pour y intégrer l'ADN viral. Le choix du site d'intégration n'est pas aléatoire et dépend de nombreux paramètres (interactions avec des protéines cellulaires, environnement du pore nucléaire, structure de la chromatine...). L'intégration est catalysée par une structure bien particulière : l'intasome, constitué d'un oligomère d'IN, de l'ADN viral ainsi que diverses protéines virales et cellulaires, bien que l'ensemble de ses constituants ne soit pas connu avec précision. Dans le cas des lentivirus, l'intégration se fait dans des zones de transcription active, notamment grâce à l'interaction du CPI avec des cofacteurs d'intégration qui vont le diriger dans des régions précises de la chromatine. L'intégration rétrovirale génère des discontinuités (gap) entre l'ADN viral et cellulaire, qui doivent être réparées pour que la transcription du provirus puisse se faire. Cette étape de réparation et les facteurs impliqués sont encore mal connus.

- **Les étapes tardives**

Après l'intégration, le virus va utiliser la machinerie cellulaire afin de permettre la production des ARN messagers qui seront ensuite traduits en protéines et précurseurs viraux pour assurer la production de nouveaux virus.

Transcription du provirus et traduction des ARNm

La transcription de l'ADN viral débute à partir de la région U3 située dans le LTR, ce dernier est un promoteur viral et permet le recrutement de facteurs de transcription. En amont de ce site d'initiation se trouvent des régions régulatrices comme des TATA box, permettant le recrutement efficace du complexe ARN polymérase II⁵⁶. Ainsi, des ARNm viraux plus ou moins épissés sont produits et exportés du noyau (avant que Tat ne soit produite, l'efficacité de transcription n'est pas très importante, mais une fois Tat fixée sur les nouveaux transcrits grâce aux régions TAR⁵⁷, ce processus va être grandement stimulé). Les ARNm non ou mono épissés requièrent la fixation de Rev sur les séquences RRE pour permettre leur exportation hors du noyau^{30,57}. Les ARNm viraux sont classés en trois catégories en fonction de leur degré d'épissage, et ne codent pas pour les mêmes protéines : les ARNm non épissés

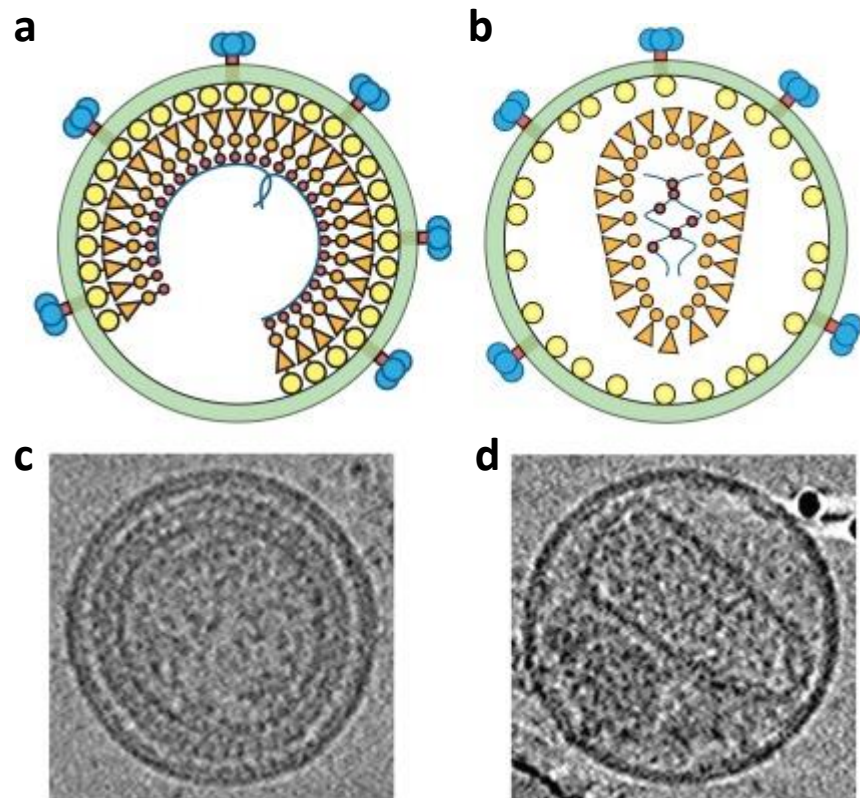


Figure 12 : Modèles et images de reconstructions tomographiques de Cryo-microscopie électronique d'une particule virale immature (c) et mature (d). La maturation d'une particule virale peut-être observée par microscopie électronique par la présence du core de capside en forme de cône (d) visible uniquement après maturation par la protéase virale⁵⁸.

vont coder pour les précurseurs polyprotéiques Gag et Gag-Pol (ils serviront aussi d'ARN génomique qui seront encapsidés dans les virions⁵⁹) ; les transcrits mono épissés coderont pour les protéines Env, Vif et Vpu ; et pour finir les ARNm multi épissés vont coder pour les protéines Rev, Tat, Vpr et Nef.

Assemblage des virions et bourgeonnement

L'assemblage des virions se fait au niveau de la membrane plasmique (MP) cellulaire. Le précurseur Gag possède un domaine d'interaction avec la MP qui va déclencher le bourgeonnement du virion après que Gag se soit oligomérisé⁶⁰. L'ARNv est ensuite recruté grâce à des interactions entre autres avec la NC et une séquence signal pour l'encapsidation présente sur cet ARN. La libération des virions immatures de la MP est assurée par le complexe cellulaire ESCRT⁶¹ (endosomal sorting complexes required for transport).

Maturation des particules virales

Une fois détachés de la MP, les virions nécessitent une dernière étape pour devenir infectieux. La PR virale va cliver les précurseurs Gag et Gag-Pol sur dix sites différents, générant les protéines MA, CA, NC, p6, PR, RT, et IN fonctionnelles^{62,63}. Au cours de ces clivages, les protéines structurales vont subir des réarrangements créant ainsi des virus matures infectieux, avec le caractéristique core viral de forme conique⁵⁸ (**Figure 12**).

Des virus exprimant une PR inactive sont déficients pour cette étape de maturation, démontrant que des protéases cellulaires ne peuvent pas assurer cette étape, mettant ainsi en avant le rôle crucial de la PR virale durant la maturation des virions.

1.3 Stratégies antirétrovirales

C'est dans les années 90 que les premières thérapies anti-VIH vraiment efficaces voient le jour avec l'apparition des trithérapies. L'avancée des connaissances sur le fonctionnement du virus a permis le développement de plusieurs molécules ciblant des étapes différentes du cycle viral. Malheureusement ces molécules ne sont pas très efficaces sur le long terme à cause des résistances que génère le virus contre ces inhibiteurs, et c'est en les combinant que l'arsenal anti-VIH commence à voir le jour. Mais encore aujourd'hui, malgré les armes dont nous disposons, il n'est toujours pas possible de guérir du SIDA et la personne infectée devra suivre un traitement à vie. D'où l'importance de sensibiliser la population quant aux risques et aux conséquences qu'être infecté implique, car le meilleur moyen pour éradiquer cette maladie est de limiter la transmission du virus.

a) Prévention

Pour une prévention efficace contre l'infection au VIH, il faut déjà savoir si on est porteur du virus. Il est estimé que plusieurs dizaines de milliers de personnes en France ignorent qu'elles sont séropositives ; ces personnes ne sachant pas qu'elles sont à risque, ne prennent ainsi pas les précautions nécessaires pour éviter la transmission du virus. Il est donc important de se faire dépister si on a le moindre doute quant à son statut sérologique.

C'est la raison pour laquelle le port du préservatif est conseillé s'il y a risque de transmission du VIH mais aussi d'autres maladies sexuellement transmissibles.

Un autre moyen de prévention a fait son apparition récemment : la prophylaxie pré-exposition⁶⁴ (PrEP). Cette stratégie est basée sur la prise d'un médicament antirétroviral qui permet à une personne séronégative à haut risque d'acquisition du VIH de réduire ce risque. Il n'y a pour le moment qu'un seul médicament utilisé pour la PrEP : le Truvada, une association de deux molécules anti-VIH (ciblant la RT), autorisée en France depuis Novembre 2015. Malheureusement, cette méthode de prévention n'est pas accessible pour tout le monde à cause de son coût, il est donc nécessaire de trouver d'autres stratégies en complément de celle-ci si l'on veut parvenir à l'éradication du virus. Et même si les méthodes thérapeutiques deviennent de plus en plus efficaces, il y a toujours des populations n'y ayant pas accès.

b) Traitements, résistance et réservoir

Il existe plusieurs raisons pour lesquelles il est toujours impossible de guérir du SIDA. Lors d'un cycle de réplication, plusieurs milliers de virus peuvent être produits par l'infection d'un seul virus pour une seule cellule. Ainsi il est estimé que plusieurs milliards de virus peuvent être produits par jour chez une personne infectée qui n'est pas sous traitement. Comme dit précédemment, lors de l'étape de transcription inverse, le virus peut générer des erreurs dans la séquence ADN reverse transcrite. Ces mutations peuvent être létales pour le virus (changement dans le cadre de lecture, protéines inactives, insertion de codons stop et autres). Mais ces mutations peuvent aussi être bénéfiques pour le virus, en lui conférant de nouvelles propriétés. Dans ce cadre, un virus résistant à un certain composé antiviral va alors se répliquer et créer une nouvelle population de virus, tous résistants à ce composé. C'est la raison pour laquelle les trithérapies ont été aussi efficaces, car en utilisant des composés ciblant plusieurs étapes différentes du cycle de réplication, cela oblige le virus à muter pour devenir résistant à chacun de ces composés pour continuer à proliférer.

Une autre raison pour laquelle il est difficile de se débarrasser du VIH est sa capacité à s'intégrer de manière latente dans certaines cellules. Ces cellules sont appelées réservoirs, elles contiennent des provirus, mais ne le répliquent pas. Ainsi elles passent inaperçues aux yeux du système immunitaire et les thérapies actuelles sont inefficaces contre ces réservoirs,

car il n'est actuellement pas possible d'éliminer le provirus de la cellule une fois qu'il est intégré. C'est une des raisons pour lesquelles il faut traiter les patients à vie, car sinon une fois la cellule activée, elle va produire du virus qui n'aura aucune contrainte pour se répliquer efficacement. En revanche, plusieurs stratégies sont à l'étude pour faire face à ce problème, comme le « kick and kill », visant à activer la cellule (kick) pour qu'elle produise du virus qui sera ensuite éliminé par le système immunitaire et accessible aux antirétroviraux⁶⁵ (kill). Récemment, l'identification de certains marqueurs de ces cellules réservoirs⁶⁶ a ouvert des portes pour le développement de composés visant à détruire ces cellules sans avoir besoin qu'elles soient réactivées. Enfin, la technologie Crispr Cas9, capable d'induire des cassures double brins au niveau de séquences bien spécifiques afin d'exciser le génome viral de l'hôte, semble très prometteuse pour le futur même si ce système d'édition génétique est à utiliser avec précaution et soulève des questions d'éthique.

La thérapie actuelle anti-VIH est basée sur l'utilisation de plusieurs molécules qui ciblent des étapes différentes du cycle de réplication du virus. Ainsi, il existe des inhibiteurs de fusion, d'entrée, de transcription inverse, d'intégration et de maturation.

- **Les inhibiteurs de fusion**

L'enfuvirtide est le seul inhibiteur de fusion utilisé en thérapie actuellement. Ce composé est un peptide qui mime le domaine HR2 de la gp41 qui va pouvoir se fixer sur le domaine HR1, inhibant ainsi le changement conformationnel de cette dernière. Sans réarrangement de la gp41, le processus de fusion ne peut pas se faire, ce qui empêche l'entrée du virus dans la cellule hôte⁶⁷.

- **Les inhibiteurs d'entrée**

Le Maraviroc (approuvé en 2007), interagit avec le corécepteur CCR5, prévenant ainsi l'interaction entre la gp120 et le CCR5 requis pour l'entrée. Seulement, cet inhibiteur n'est actif que contre les souches virales à tropisme CCR5, car dans le cas où le virus possède un tropisme CXCR4, il n'a pas besoin du CCR5 pour entrer dans la cellule cible, le Maraviroc est par conséquent inactif contre ces virus⁶⁸.

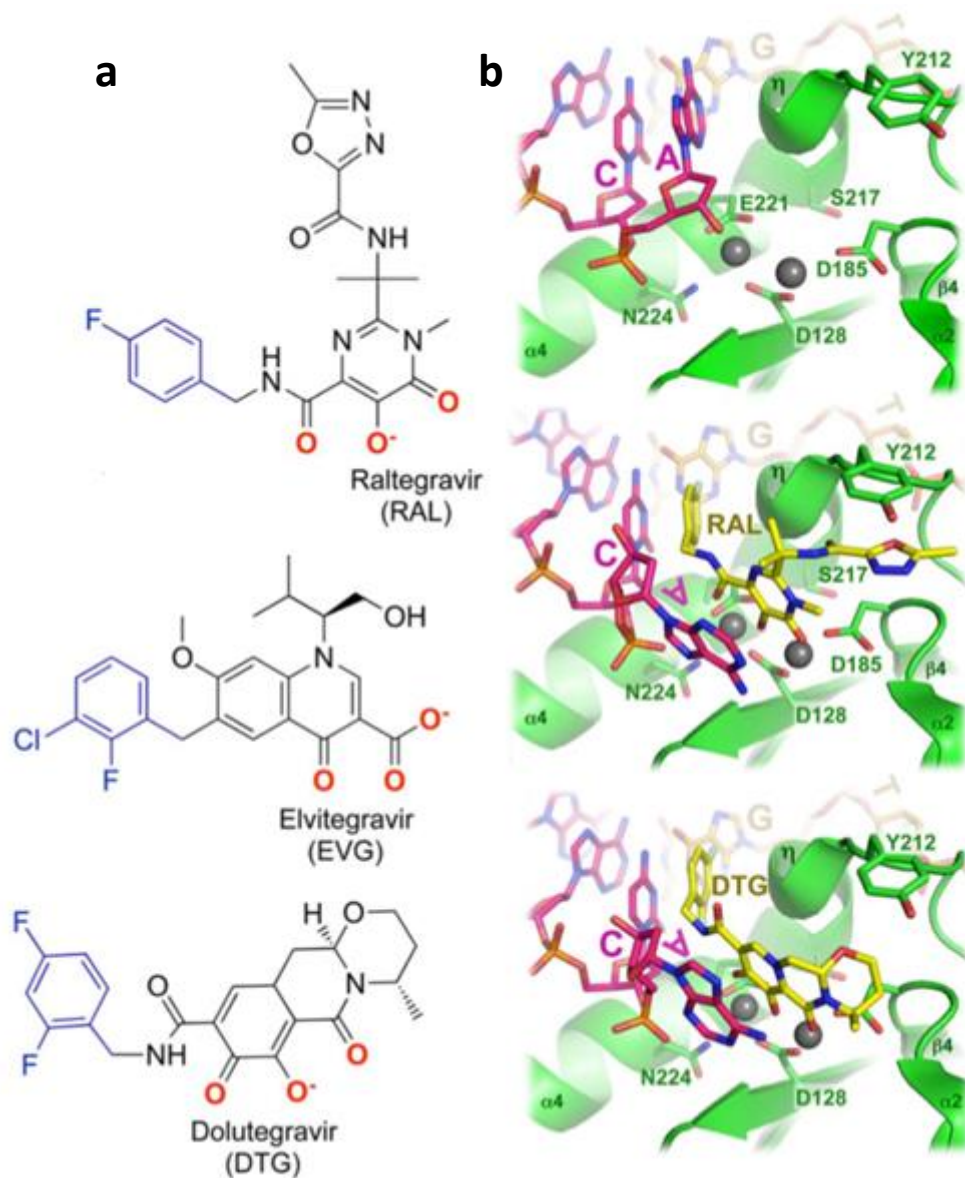


Figure 13 : Inhibiteurs de l'activité de transfert de brin de l'IN du VIH-1. (a) Structures chimiques des INSTIs utilisés en clinique (Raltegravir, Elvitegravir et le Dolutegravir). **(b)** Site actif de l'intasome de PFV avant (image du haut) et après (image du milieu) la fixation du Raltegravir ou du Dolutegravir (image du bas)⁶⁹.

- **Les inhibiteurs de transcription inverse**

Il existe deux types d'inhibiteurs de la transcription inverse : les NRTIs (nucleoside reverse transcriptase inhibitors) et les NNRTIs (non nucleoside reverse transcriptase inhibitors). Les NRTIs sont des analogues de nucléosides et de nucléotides terminateurs de chaîne. Une fois dans la cellule, ils sont phosphorylés par des kinases cellulaires ce qui active ces composés. Leur particularité est qu'ils ne possèdent pas de groupement 3'OH, ainsi s'ils sont incorporés, le prochain nucléotide ne pourra pas former de liaison phosphodiester avec le suivant stoppant par conséquent la synthèse de l'ADN viral⁷⁰. Parmi ces NRTIs on retrouve la Zidovudine (première molécule anti-VIH approuvée par la FDA en 1985) et le plus utilisé aujourd'hui : le Ténofovir (approuvé en 2001).

Le second type d'inhibiteurs, les NNRTIs, cible directement la RT en interagissant dans un domaine proche du site catalytique au niveau de la sous-unité p66. Cette interaction induit un changement conformationnel de la RT la rendant inactive⁷¹. Parmi les NNRTIs, on retrouve l'Efavirenz, la Nevirapine et l'Etravirine.

- **Les inhibiteurs d'intégration**

Tous les inhibiteurs d'intégration actuels utilisés en thérapie partagent le même mécanisme d'action sur l'IN : ce sont des inhibiteurs qui touchent spécifiquement l'étape de transfert de brin, ils sont appelés INSTIs (IN strand transfer inhibitors). Ils se lient au niveau du site catalytique de l'enzyme quand elle est en interaction avec l'ADN viral (**Figure 13**), empêchant ainsi l'ADN cellulaire de pouvoir se positionner au bon endroit dans le site catalytique de l'enzyme⁷². Ces inhibiteurs partagent également certaines fonctionnalités : la chélation d'ions Mg^{2+} grâce à une triade d'atomes d'oxygène, ces ions sont normalement requis pour l'étape d'intégration. Une autre force de ces inhibiteurs est leur capacité à déplacer l'extrémité 3' de l'ADN viral traité grâce à une de leur chaîne aromatique, ainsi l'extrémité de l'ADN viral qui sert normalement à faire la réaction d'intégration se retrouve complètement hors du site catalytique. Toutes ces caractéristiques en font de puissants inhibiteurs, et il se trouve qu'ils sont également actifs contre des IN d'autres rétrovirus car le site actif est très conservé entre les différentes enzymes⁷³. Mais ils ont tout de même une faiblesse, en se fixant tous dans la même région de l'IN, l'apparition de virus résistants à un des inhibiteurs entraîne souvent une résistance croisée avec les autres.

Le premier inhibiteur d'intégration approuvé contre le SIDA a été le Raltegravir en 2007 puis l'Elvitegravir (2012) et le Dolutegravir (2013) ont fait leur apparition sur le marché. Actuellement, deux autres inhibiteurs sont en phase clinique : le Bictegravir et le Cabotegravir.

Pour pallier aux problèmes de résistances contre les INSTIs, de nouvelles stratégies basées sur le développement de composés ne ciblant pas le site catalytique de l'IN, mais des protéines cellulaires importantes pour l'intégration ont l'avantage d'être théoriquement actifs

contre les souches de virus résistantes aux INSTIs. Parmi les inhibiteurs non catalytiques décrits actuellement, les ALLINIs (allosteric IN inhibitors) aussi appelés LEDGINs (LEDGF-IN interaction site) semblent être les plus prometteurs. Ils ont initialement été développés dans le but d'empêcher l'interaction entre l'IN et son principal cofacteur d'intégration LEDGF⁷⁴. Bien qu'étant de puissants inhibiteurs d'intégration, il se trouve que ces composés n'ont en réalité pas le mécanisme d'action attendu : l'assemblage de l'intasome du VIH est très sensible à l'oligomérisation de l'IN, afin de s'organiser correctement sur l'ADN viral ; les ALLINIs se fixent au niveau d'interfaces d'oligomérisation de l'IN, induisant ainsi la multimérisation anarchique de l'enzyme et la formation d'agrégats⁷⁵⁻⁷⁷. Ces agrégats ne sont plus capables d'interagir correctement avec l'ADN viral pour former des complexes IN-ADN fonctionnels. Cependant, les ALLINIs sont également considérés comme des inhibiteurs de la maturation des virions, car l'agrégation de l'IN dans les virions conduit à des défauts de maturation^{78,79}.

- **Les inhibiteurs de protéase**

Les inhibiteurs de protéase sont des analogues non clivables de peptides qui miment le substrat de la protéase. Ainsi, la protéase va reconnaître et se fixer sur ces peptides au lieu des précurseurs polyprotéiques, qui ne seront alors plus clivés. Ceci bloque l'étape de maturation des virions qui ne deviendront alors pas infectieux⁸⁰. Il y a aujourd'hui dix inhibiteurs de protéase approuvés par la FDA, comme que le Saquinavir (1^{er} inhibiteur de protéase approuvé en 1995) ou le Darunavir (le dernier en date, approuvé en 2006). Les inhibiteurs de protéase sont très efficaces et jouent un rôle très important dans les trithérapies, malheureusement ils entraînent des complications pour les patients et sont souvent toxiques, c'est la raison pour laquelle d'autres composés de la même famille sont à l'étude dans le but d'identifier des inhibiteurs qui soient mieux tolérés par les patients.

c) Essais Vaccinaux

Les vaccins thérapeutiques reposent sur deux stratégies pour venir à bout du VIH. La première est de faire sortir les virus des cellules réservoirs qui ne sont pas accessibles par les trithérapies, pour ensuite les cibler avec un vaccin. La deuxième consiste à stimuler la réponse immunitaire pour qu'elle soit assez puissante pour pouvoir contrôler la réplication virale en absence de traitement. A l'heure actuelle aucune de ces deux stratégies n'est applicable à cause de nombreuses contraintes telles que la variabilité du virus et la difficulté à reconnaître des cellules infectées de manière latente. Malgré les centaines d'essais vaccinaux menés depuis la découverte du VIH, un seul (RV144 débuté en 2003) a pu montrer un effet significatif sur l'inhibition de la réplication virale. Cet essai a été conduit en Thaïlande sur près de 16 000 volontaires, et a montré une efficacité proche de 31%⁸¹. Il faudra donc attendre quelques années pour espérer entrevoir une solution efficace utilisable à grande échelle.

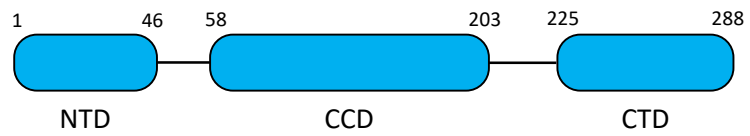


Figure 14 : Organisation schématique de l'IN du VIH-1. L'IN du VIH-1 est une protéine de 288 acides aminés organisée en trois domaines (NTD : domaine N-terminal ; CCD : domaine catalytique ; CTD : domaine C-terminal).

2 L'intégration du VIH-1

2.1 Structure de l'IN du VIH-1

Après maturation du précurseur Gag-Pol par la PR dans les virions, l'IN est alors produite sous forme d'une protéine de 288 acides aminés (soit 32 kDa). Elle est composée de trois domaines (**Figure 14**) : le domaine N-terminal (NTD), le corps catalytique (CCD) et le domaine C-terminal (CTD)⁸². Chacun de ces domaines est requis pour l'intégration. Pour le moment, aucune structure tridimensionnelle de la protéine entière n'a pu être résolue, limitant énormément la rationalisation de drogues contre cette protéine. En revanche, la structure des domaines isolés et deux à deux a pu être résolue ainsi que celle de plusieurs intasomes rétroviraux, ce qui a servi de base pour la compréhension du comportement de l'IN.

a) Le domaine N-terminal (NTD)

La structure du NTD seul a été résolue par RMN, ce qui a permis de mettre en évidence un motif HHCC proche d'un domaine à doigt de zinc pouvant fixer un ion Zn^{2+} , nécessaire à la structuration du NTD isolé⁸³.

b) Le corps catalytique (CCD)

La structure cristallographique du CCD a été obtenue par introduction de la mutation F185K augmentant significativement la solubilité de ce domaine tout en conservant l'activité de l'enzyme^{84,85}. Le domaine CCD contient le site actif de l'IN, composé de la triade catalytique DDE pouvant fixer un cofacteur métallique (Mn^{2+} ou Mg^{2+}) requis pour l'activité enzymatique. Ce motif est très conservé parmi les intégrases de rétrotransposons et de rétrovirus, la mutation d'un résidu de cette triade abolie l'activité de ces enzymes.

c) Le domaine C-terminal (CTD)

La structure du CTD a été résolue par RMN, identifiant un repliement proche d'un domaine SH3 (sarc homology 3 domain) connu pour avoir de l'affinité pour l'ADN⁸⁶.

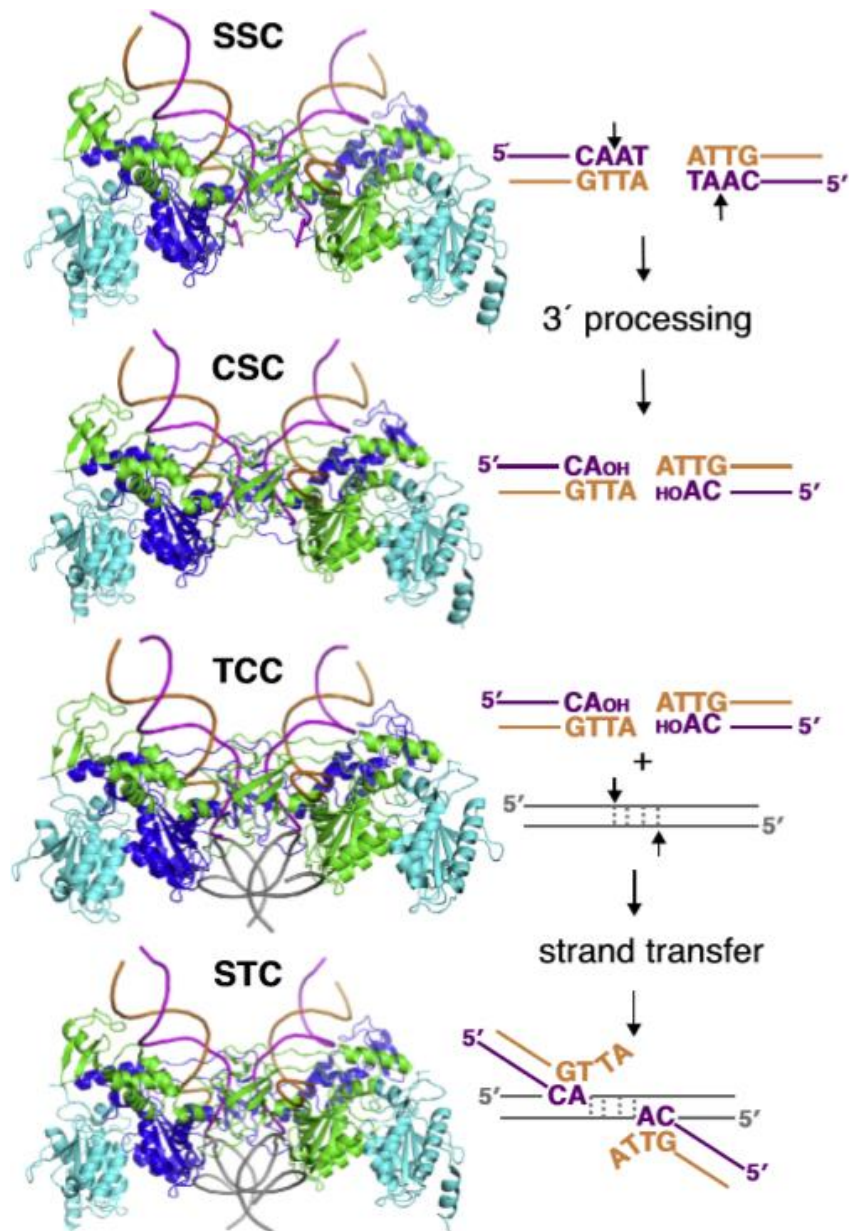


Figure 15 : Les différents états de l'intasome de PFV. Les différentes étapes qui conduisent à la formation du SSC (Stable Synaptic Complex) ne sont pas encore bien décrites. Le SSC est mûré (3' processing) devenant le CSC (Cleaved Synaptic Complex). Après reconnaissance de l'ADN cible il devient le TCC (Target Capture Complex) puis finalement le STC (Strand tranfert Complex) qui catalyse l'intégration. En bleu et vert sont montrés les protomères d'IN interagissant avec l'ADN cible. En cyan les IN qui ne sont pas impliquées directement dans l'association avec l'ADN cible. Les extrémités de l'ADN viral sont en violet et orange et l'ADN cible en gris⁸⁷.

Pendant des années il a généralement été admis que les domaines isolés pouvaient avoir des fonctions bien distinctes les uns des autres (oligomérisation, interaction avec l'ADN viral ou cellulaire, etc...). Mais les avancées récentes dans la résolution des structures d'intasomes de différents rétrovirus, ont montré une organisation bien plus complexe que celle de l'intasome de PFV qui a longtemps servi de base structurale^{34,72}. En effet, comparé à l'intasome de PFV qui est constitué d'un tétramère d'IN, l'intasome de RSV⁸⁸ et MMTV⁸⁹ est composé d'un octamère d'IN, alors que l'intasome de MVV⁹⁰ (un lentivirus comme VIH) est un hexadodécamère. Ces structures ont non seulement permis de mettre en évidence des différences et des similarités entre les intasomes rétroviraux, mais aussi de voir qu'en fonction de sa position au sein de l'intasome, chaque domaine CTD, CCD ou NTD pouvait avoir un rôle différent, révélant ainsi la pléiotropie de chacun de ces domaines.

d) Les intasomes rétroviraux

Comme dit précédemment, l'intégration rétrovirale est catalysée au sein d'un complexe nucléoprotéique : l'intasome. L'assemblage de l'intasome se fait progressivement au cours du cycle de réplication du virus⁸⁷ (**Figure 15**). Suite à la transcription inverse, il y a dans un premier temps la formation du SSC (stable synaptic complex) qui permet la maturation des extrémités de l'ADN viral (3' processing) qui devient alors le CSC (cleaved synaptic complex). Une fois en contact avec la chromatine l'intégration va finalement pouvoir être réalisée par le STC (strand transfer complex). L'assemblage de l'intasome est mal connu : l'état oligomérique, les partenaires, la structure et la chronologie précise de chacune de ces étapes restent encore à être précisément déterminés.

Le premier intasome (CSC) à avoir été caractérisé structuralement est celui du spumavirus PFV en 2010⁷², par diffraction aux rayons X, révélant pour la première fois la fonctionnalité de l'intasome en tant que multimère d'IN et élucidant également le mécanisme d'action des inhibiteurs de transfert de brin. L'architecture tétramérique de l'intasome de PFV a été longtemps considérée comme une composante universelle de la machinerie d'intégration rétrovirale. Cependant, ces dernières années, la résolution de la structure d'intasomes de quatre rétrovirus différents a permis de réfuter cette hypothèse, révélant des architectures bien plus complexes (**Figure 16**).

Alors que l'intasome de PFV est constitué de quatre molécules d'IN, les intasomes de MMTV (bétarétrovirus) et RSV (alpharétrovirus) en ont huit. Celui de MVV est encore plus gros avec seize IN. Malgré les différences architecturales de ces structures, il semble y avoir une caractéristique qui les relie : la présence d'un « cœur » ressemblant énormément à l'intasome de PFV (schématisé en couleur sur la **Figure 16**), appelé par la suite CIC⁸⁷ (conserved intasome core). Ces avancées permettront de mieux appréhender les résistances aux inhibiteurs d'intégration pour en désigner de meilleurs, et de comprendre pourquoi il y a une telle hétérogénéité parmi ces structures en comprenant comment les rétrovirus tirent avantage d'avoir des intasomes de tailles différentes.

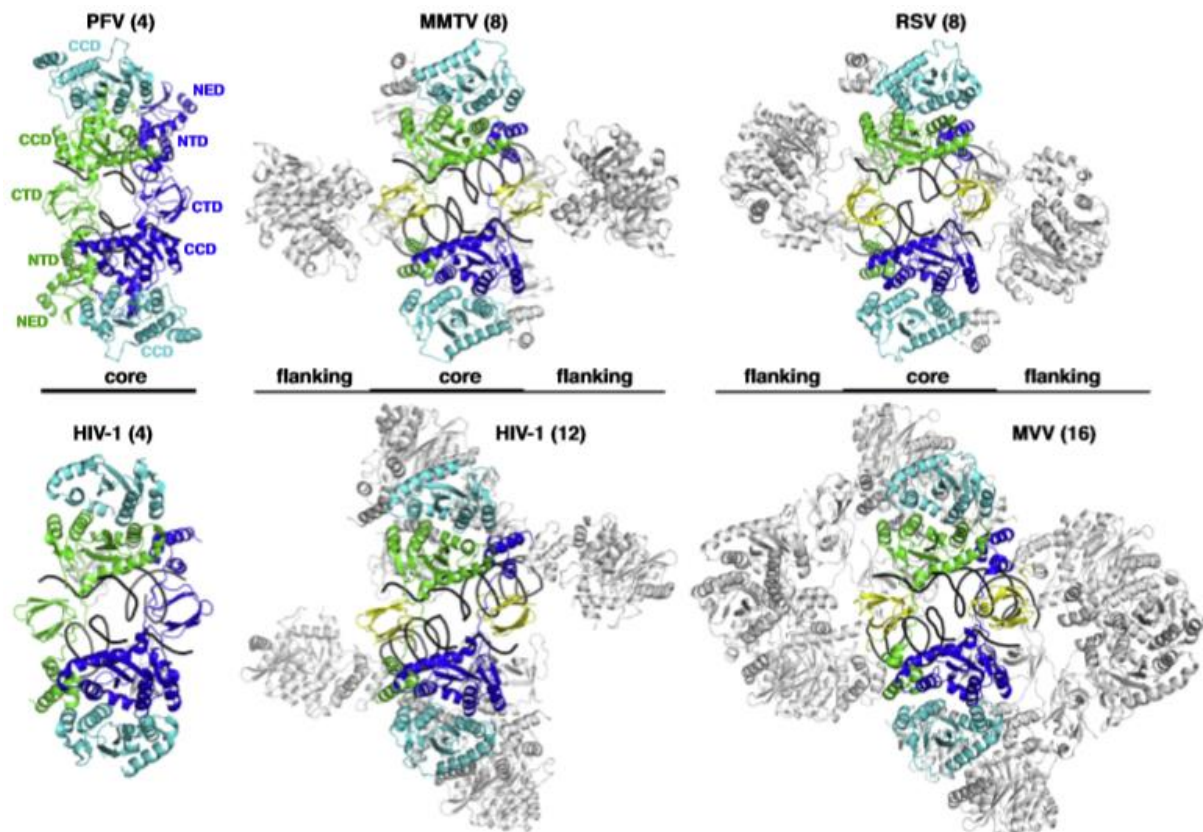


Figure 16 : Structures de différents intasomes rétroviraux. Les IN en couleur font partie du CIC (Conserved Intasome Core). PFV CSC (pdb 3oy9), MMTV CSC (pdb 3jca), RSV STC (pdb 5ejk), Tetramère du VIH-1 STC obtenu avec l'IN fusionnée au domaine SsO7d (pdb 5u1c) et MVV CSC (pdb 5m0q)⁸⁷.

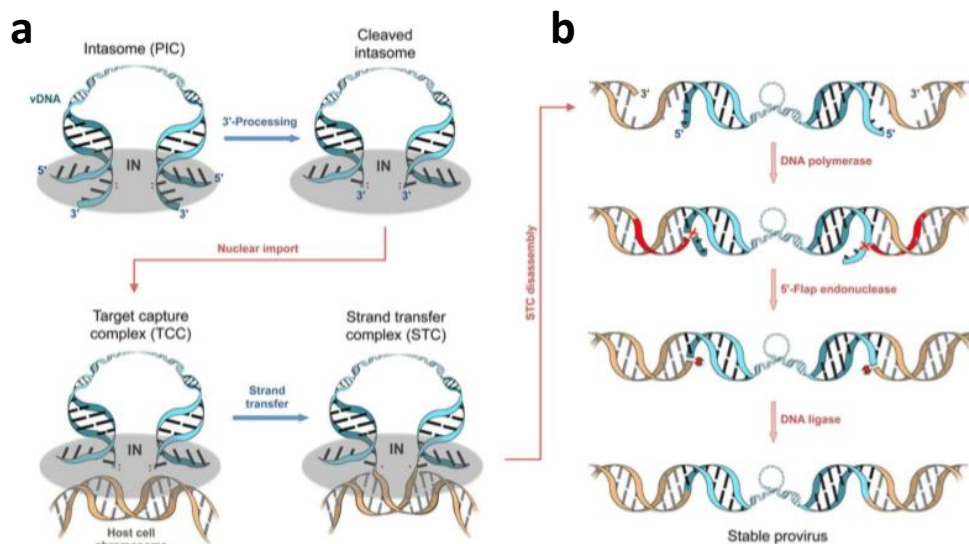


Figure 17 : Les étapes de l'intégration rétrovirale. Réactions catalysées par l'IN (a) et par des enzymes cellulaires (b). (a) L'intasome (en gris) est mûri par clivage d'un dinucléotide aux extrémités 3' de l'ADN viral. Après l'import nucléaire et reconnaissance de l'ADN cible, les extrémités 3' de l'ADN viral sont insérées dans l'ADN de l'hôte lors du transfert de brin. (b) L'intégration est finalisée par la réparation des discontinuités aux extrémités de l'ADN viral par des enzymes cellulaires aboutissant à l'insertion complète et stable du provirus⁶⁹.

Pour ce qui est du VIH-1, la faible solubilité de l'enzyme et le fait qu'elle soit presque inactive sans l'addition de cofacteurs comme LEDGF, ont été de grandes limites pour l'obtention d'une structure que ce soit pour l'intasome ou la protéine seule. Pour pallier à ces problèmes, une étude a été menée sur une intégrase fusionnée avec Sso7d, une protéine augmentant à la fois sa solubilité et son activité⁹¹. L'intasome résultant de cette fusion Sso7d-IN a pu être cristallisé, révélant un intasome tétramérique ainsi que d'autres espèces de plus haut poids moléculaire, avec toujours le même CIC que dans les autres intasomes⁹². De plus, l'addition d'un fragment de LEDGF (l'IBD pour intégrase binding domain) à Sso7d-IN a montré l'apparition d'une nouvelle population constituée majoritairement de dodécamères (douze IN). Ces résultats, bien que très encourageants, sont à prendre avec précaution car il n'est pas exclu que le domaine de fusion Sso7d altère la configuration de l'intasome du VIH-1 telle qu'elle serait normalement attendue avec une enzyme sauvage.

2.2 Mécanisme d'action de l'IN

Dans la cellule, l'intégration se fait en trois étapes consécutives. Il y a dans un premier temps le 3' processing qui consiste en la maturation des extrémités de l'ADN viral par clivage de deux nucléotides à chacune des extrémités 3'. Ceci permet ensuite l'intégration de ces extrémités dans l'ADN cellulaire : c'est l'étape de transfert de brin. La dernière étape n'est pas catalysée par l'IN, c'est la réparation des gaps au niveau du site d'intégration. Le 3' processing ainsi que le transfert de brin peuvent être reproduits *in vitro*.

a) Le 3' processing

Pour former le SSC, l'IN interagit spécifiquement avec une séquence ADN précise située à chaque extrémité du LTR formant ainsi ce complexe synaptique (**Figure 17**). Le processing de l'ADN viral a lieu dans le cytoplasme après la transcription inverse⁹³ et nécessite un cofacteur métallique (Mg^{2+} ou Mn^{2+}). L'IN va hydrolyser une liaison phosphodiester libérant un dinucléotide GT à chacune des extrémités 3' de l'ADN, générant ainsi une fonction hydroxyle (OH) libre⁹⁴. Cette configuration devient alors compatible pour l'intégration.

b) Le transfert de brin

Après l'import nucléaire du CSC et interaction avec la chromatine, l'IN va catalyser la seconde étape du processus d'intégration au sein du TCC. Les groupements hydroxyles en 3' de l'ADN viral générés lors du processing vont servir de substrats nucléophiles pour couper une liaison phosphodiester de l'ADN cellulaire. Cette réaction nécessite également un cofacteur métallique (Mg^{2+} ou Mn^{2+}) et aboutit à un clivage des deux brins de l'ADN cible et la liaison des deux extrémités de l'ADN viral dans la chromatine. Ceci génère un décalage de

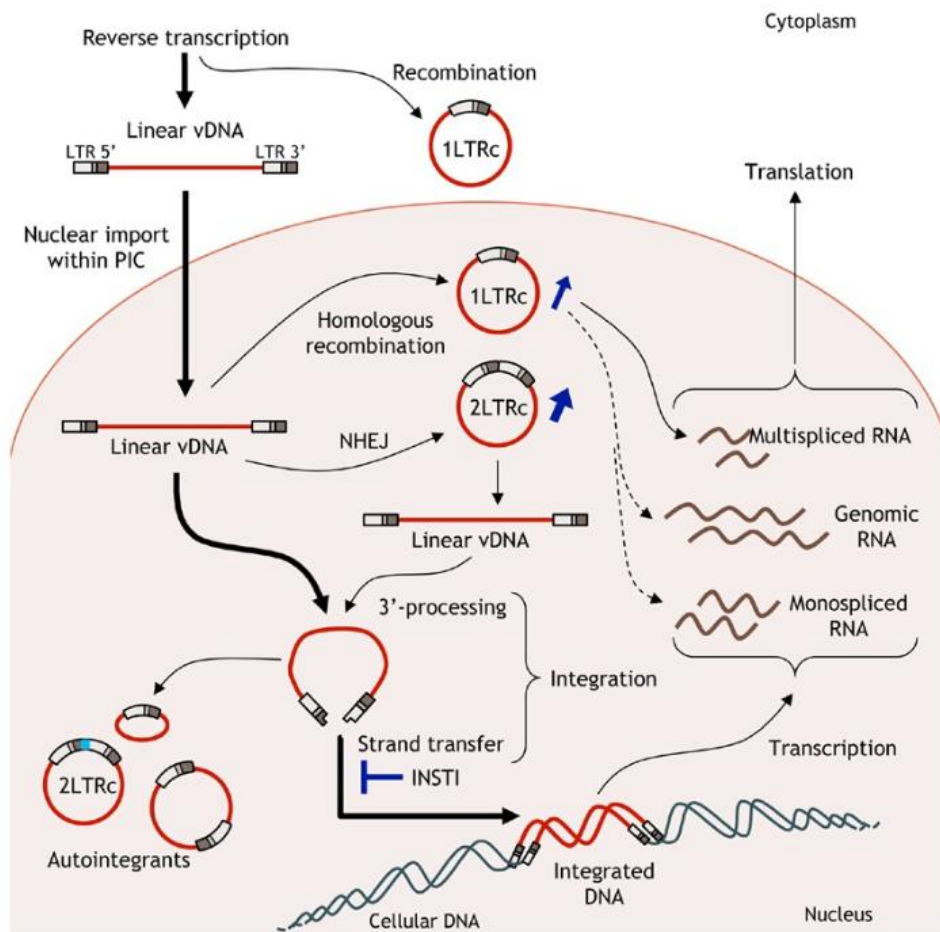


Figure 18 : Devenir des formes non intégrées de l'ADN viral. L'ADN viral linéaire provenant de la transcription inverse peut avoir plusieurs devenir s'il n'est pas intégré (cercles à 1 ou 2 LTR ou différentes formes résultant de l'autointégration). Les flèches bleues indiquent que l'utilisation d'inhibiteurs d'intégration augmentent la proportion des formes circulaires⁹⁵.

cinq paires de base dans l'ADN cible pour ce qui est du VIH, qui doit ensuite être réparé pour que l'intégration soit finalisée.

c) Réparation post-intégration

Afin d'assurer la continuité de l'ADN, le décalage (gap) entre l'extrémité 5' de l'ADN viral et 3' de l'ADN cellulaire résultant de l'intégration doit être réparé. Cette étape est réalisée par des enzymes cellulaires impliquées dans les voies de réparation de l'ADN telles que FEN-1, mais ce processus et les facteurs intervenant dans la réparation de ces gaps ne sont pas encore totalement élucidés. Cependant plusieurs candidats sont suspectés d'être impliqués dans cette voie de réparation⁹⁶ :

La voie NHEJ (non homologous end-joining) qui est un mécanisme de réparation des cassures d'ADN.

La ligase IV joint les extrémités d'ADN au niveau de cassures double brins lors de réparations non homologues.

RAD51, une recombinaise cellulaire, pourrait également jouer un rôle dans les étapes de la réparation de l'ADN⁹⁷.

d) Les produits non intégrés

Il existe aussi des formes d'ADN non intégrées (**Figure 18**) retrouvées principalement dans le noyau de la cellule. Ces différentes formes peuvent être originaires de plusieurs sources (mauvaise transcription inverse, autointégration, traitement par les INSTIs/ALLINIs, etc...). Parmi ces formes, on retrouve des ADN linéaires et circulaires à 1 ou 2 LTR. Les cercles à 2 LTR sont exclusivement retrouvés dans le noyau et sont ainsi utilisés comme marqueurs de l'import nucléaire⁹⁸. Récemment, il a été montré que les cercles à 2 LTR pouvaient servir de substrat pour l'intégration, et que certains produits non intégrés étaient capables d'exprimer des protéines virales. Ces résultats indiquent que ces formes d'ADN non intégrés peuvent être considérés comme des réservoirs et semblent être impliqués dans la latence⁹⁵.

2.3 L'intégration *in vitro*

Il est possible de reproduire l'activité de l'IN *in vitro* grâce à différents tests qui ont pu être optimisés pendant des années. Du test de processing et de transfert de brin au test d'intégration concertée, il est maintenant possible de faire de l'intégration *in vitro* avec des intasomes purifiés. Ces différentes techniques nous permettent de mieux appréhender le comportement des IN et sont des outils puissants pour tester et développer de potentiels

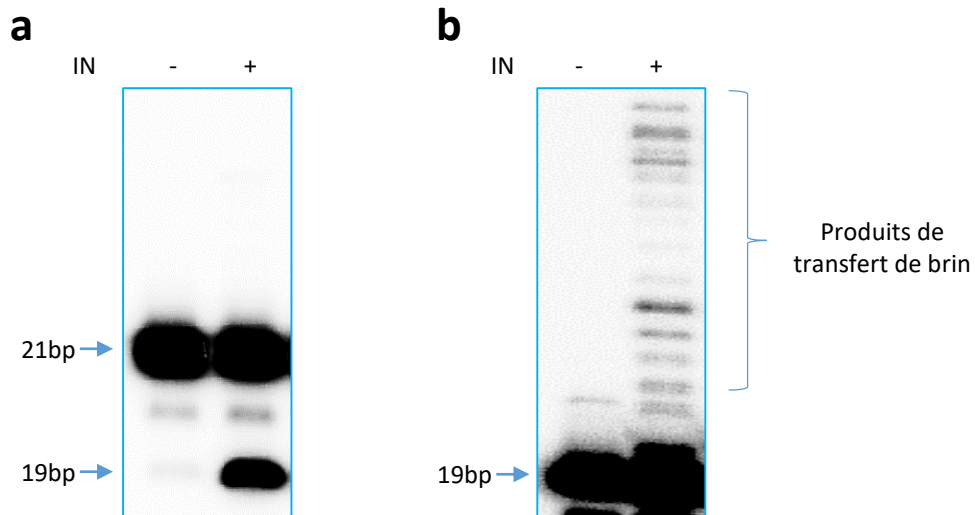


Figure 19 : Profils des tests d'activité catalytique de 3' processing (a) et de transfert de brin (b) de l'IN du VIH-1. L'ADN est marqué radioactivement en 5'. Les produits sont séparés sur gel de polyacrylamide et révélés par autoradiographie.

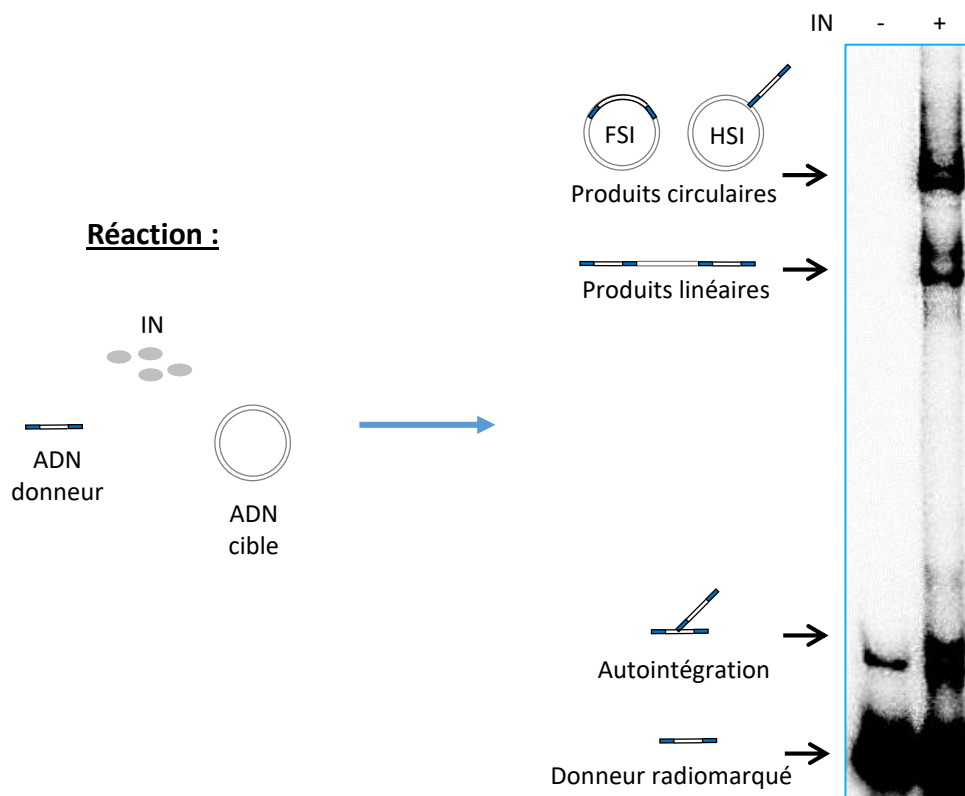


Figure 20 : Profil du test d'intégration concertée de l'IN du VIH-1. L'ADN donneur est marqué radioactivement en 5'. Les différents produits d'intégration sont séparés sur gel d'agarose et visualisés par autoradiographie. (FSI : Full Site Integration ; HSI : Half Site Integration).

nouveaux inhibiteurs.

a) Test de processing et de transfert de brin

Dans le cas du VIH, le test de processing utilise un oligonucléotide (ODN) double brin de 21pb mimant l'extrémité U5 du LTR (séquence reconnue spécifiquement par l'IN) comme substrat. Cet ODN est marqué radioactivement au P³² en 5', ainsi quand l'IN aura hydrolysée le dinucléotide GT de l'extrémité 3', le produit de 19pb sera visualisable sur gel de polyacrylamide dénaturant (**Figure 19a**).

L'étape de transfert de brin peut être étudiée indépendamment de l'activité de processing. Pour cela un ODN déjà traité (un hybride 21-19pb) est utilisé comme substrat. L'IN va catalyser l'insertion d'un ODN dans un autre en réalisant l'étape de transfert de brin. Dans ce test l'insertion se fait de manière aléatoire et conduit par conséquent à l'apparition de produits d'intégration de tailles différentes (**Figure 19b**).

b) Le test d'intégration concertée

Le test d'intégration concertée (IC) permet de se rapprocher un peu plus du mécanisme d'intégration tel qu'il est attendu dans la cellule. Dans les tests de processing et de transfert de brin, seule une extrémité du substrat est prise en charge par l'IN, alors que dans la cellule les deux extrémités de l'ADN viral sont intégrées de manière concomitante. Le substrat utilisé pour les tests IC comprend les deux extrémités des séquences du LTR, ainsi l'intégrase va pouvoir prendre en charge ces deux extrémités. Les composants minimaux du test IC sont l'IN, un ADN mimant l'ADN viral (qui comprend les deux extrémités des LTR) et un ADN mimant l'ADN cible. Avec ces composants, l'IN peut catalyser différents produits d'intégration visualisables sur gel d'agarose (**Figure 20**). Parmi ces produits on distingue de l'autointégration résultant de l'insertion d'un ADN viral dans un autre ADN viral, et de l'hétérointégration résultant de l'intégration d'un ADN viral dans le plasmide. Les produits d'hétérointégration sont multiples : on distingue des formes circulaires (résultant de l'intégration d'un ADN viral) et linéaires (dû à l'intégration de deux ADN viraux dans le plasmide ce qui le circularise). Les produits linéaires sont issus d'une réaction d'intégration concertée et sont dits « full site » (FSI pour full site intégration), alors que les produits circulaires peuvent être full site ou half site (HSI) ce qui signifie qu'une seule extrémité de l'ADN viral s'est intégrée dans le plasmide.

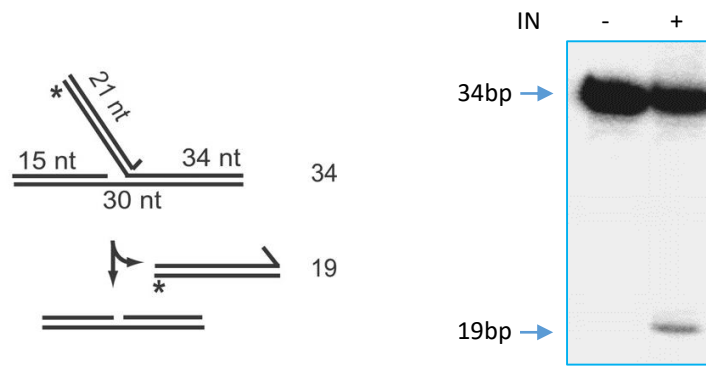


Figure 21 : Profil du test de désintégration de l'IN du VIH-1⁹⁹.

c) La désintégration

La désintégration correspond à l'étape inverse du transfert de brin et se fait uniquement *in vitro*¹⁰⁰. Elle peut être catalysée par le domaine CCD seul et nécessite la présence de Mg^{2+} ou Mn^{2+} . Pour ce test, un substrat ADN avec une structure particulière mimant un produit d'intégration partiel est utilisé (**Figure 21**) que l'IN peut cliver. Même si cette activité ne semble pas avoir de relevance *in vivo*, elle a permis de mieux comprendre le fonctionnement du site actif de l'IN.

2.4 Activités non catalytiques de l'IN

Comme soulevé précédemment, l'IN a un rôle pléiotrope en participant et en régulant d'autres étapes du cycle viral que l'intégration. Des expériences de mutagenèse ont permis de séparer les mutants d'IN présentant un défaut de réplication en deux classes en fonction de leur phénotype. Les mutants déficients pour l'étape d'intégration sont dits de classe I, alors que les mutants d'IN qui sont affectés pour une autre étape que l'intégration sont de classe II¹⁰¹. Les mutants de classe II ont été la preuve que l'IN est également importante pour d'autres étapes que l'intégration¹⁰².

a) Rôle de l'IN dans la transcription inverse

Une interaction directe entre la RT et l'IN a été démontrée *in vitro*, et des expériences de RMN entre la RT et le CTD de l'IN ont permis d'identifier des résidus clés pour cette interaction¹⁰³. La mutation de résidus de l'IN conduisant à une perte d'interaction avec la RT entraîne un déficit de réplication des virus contenant ces IN, et spécifiquement au niveau de l'étape de transcription inverse. Le rôle exact de l'IN au cours de cette étape n'est pas encore très bien connu, et les tests de transcription inverse *in vitro* ne sont pas très efficaces car la RT a une très faible processivité. Cependant, il a été observé que l'IN pouvait stimuler l'initiation et l'élongation de la transcription inverse *in vitro*, en augmentant la processivité de la RT. Ces résultats révèlent l'importance de l'interaction fonctionnelle RT-IN lors de l'étape de transcription inverse et suggèrent que l'IN pourrait jouer un rôle dans les événements précoces de la transcription inverse.

b) Rôle de l'IN dans l'import nucléaire

Etant donné que le CPI des lentivirus peut entrer de manière active dans le noyau et infecter des cellules qui ne se divisent pas, il a été proposé que l'IN pouvait jouer un rôle dans ce transport actif. En effet une séquence de localisation nucléaire (NLS) a été identifiée dans

le domaine CCD de l'IN, et des mutations au niveau de cette séquence abolissent l'import nucléaire, révélant une NLS fonctionnelle requise pour l'infection de cellules en division ou non¹⁰⁴. Une autre étude a montré que des mutations introduites dans le CTD de l'IN conduisaient à une diminution de l'import nucléaire de l'ADN viral en diminuant l'interaction de l'IN avec TRN-SR2, une protéine cellulaire impliquée dans l'import nucléaire du CPI¹⁰⁵.

c) Rôle de l'IN dans les phases tardives du cycle viral

Comme mentionné précédemment, l'utilisation des ALLINIs inhibe la réplication virale en induisant une multimérisation aberrante de l'IN dans les virions. L'utilisation de ces inhibiteurs entraîne un défaut de maturation, visible par une forme aberrante du core viral où l'IN se retrouve à l'extérieur du core⁷⁹. La manière dont l'IN participe à la morphogénèse des virions a pu être élucidée par la mise en évidence que l'IN a la capacité d'interagir avec l'ARN viral dans les virions, et que cette interaction est importante pour la maturation des particules virales. Cette étude a également permis de mettre en évidence un mécanisme d'action des ALLINIs qui est d'inhiber l'interaction IN-ARN dans les virions, conduisant ainsi à un défaut de maturation.

2.5 Les cofacteurs de l'intégration

L'intégration rétrovirale est un procédé complexe qui ne se fait pas aléatoirement dans la cellule, elle est régulée par de nombreux paramètres qui peuvent différer selon le virus. Bien qu'*in vitro* l'intégration est possible avec seulement de l'IN recombinante, il a été démontré par de nombreuses approches que des protéines cellulaires sont importantes pour l'intégration *in vivo*. Le plus documenté à ce jour est LEDGF/p75 étant donné l'importance qu'il joue au cours de l'intégration, même si d'autres cofacteurs ont été identifiés, leurs effets sur l'intégration sont moindres que LEDGF ou leur rôle n'est pas encore totalement élucidé, mais ils constituent tout de même de bonnes pistes quant au développement de composés antiviraux visant ces cofacteurs. Récemment, CPSF6 a été validé comme cofacteur de l'intégration du VIH-1 et son implication au cours du processus d'intégration semble être tout aussi important que l'est LEDGF.

Integrase interactor 1 (INI1)

INI1 est une protéine de 385 acides aminés qui fait partie du complexe de remodelage SWI/SNF. Cet activateur transcriptionnel interagit directement avec l'IN du VIH-1¹⁰⁶ et est proposé comme participant au choix du site d'intégration, car *in vitro* SWI/SNF permet de restaurer l'intégration dans une matrice chromatinisée¹⁰⁷. Il a également été montré que le domaine de INI1 interagissant avec l'IN (appelé S6), a la capacité de réduire la libération des virions quand il est surexprimé dans les cellules productrices de virus¹⁰⁸. Même si le rôle

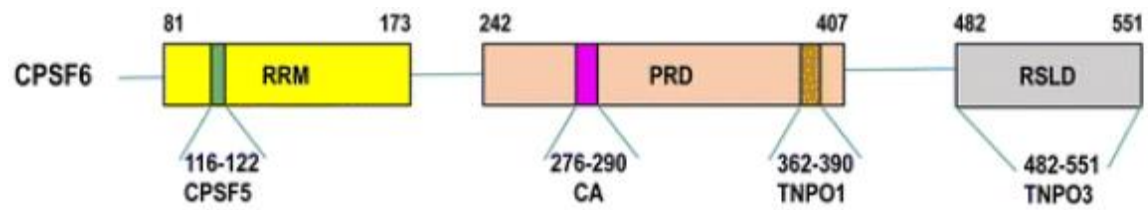


Figure 22 : Organisation schématique de CPSF6. Les domaines d'interaction de CPSF6 (RRM : RNA recognition motif ; PRD : Pro-rich domain ; RSLD : RS-like domain) avec diverses protéines sont indiqués en bas de l'image. (CA : capsid du VIH-1)¹⁰⁹.

précis de IN1 au cours du cycle de réplication n'est pas encore élucidé, il semble qu'il puisse avoir un rôle à la fois dans les phases pré et post intégration.

Barrier to autointegration factor (BAF)

BAF est une protéine de 89 acides aminés impliquée dans l'architecture nucléaire et est retrouvée dans les CPIs de cellules infectées¹¹⁰. Plusieurs rétrovirus (dont MLV et VIH) ont utilisé BAF à leur avantage pour limiter les événements d'autointégration qui sont létaux pour le virus.

High mobility group chromosomal protein A1 (HMGA1)

HMGA1 est une protéine impliquée dans la régulation de la transcription et de la structure de la chromatine. Il est proposé que HMGA1 facilite l'interaction de l'IN sur les LTR de l'ADN viral par modulation de la structure des LTR¹¹¹. Même si la protéine HMGA1 ne semble pas impliquée dans l'intégration rétrovirale, il semblerait qu'elle puisse participer à la transcription du provirus. Des sites d'interaction à HMGA1 ont été identifiés dans le LTR en 5' pouvant potentiellement participer au recrutement de facteurs de transcriptions tels que SWI/SNF, régulant ainsi la transcription du provirus¹¹².

FACT

Le FACT (Facilitates chromatin transcription) est composé des protéines SSRP1 et SPT16¹¹³, intervenant dans la réplication, la transcription et la réparation de l'ADN. Le FACT est un complexe de remodelage de la chromatine, il est capable de désassembler partiellement et de réassembler le nucléosome lors de la transcription associée à l'ARN pol II¹¹⁴. Ce complexe a été identifié au laboratoire comme cofacteur de l'intégration du VIH-1. En désassemblant partiellement le nucléosome il augmente ainsi l'accessibilité de ce dernier pour l'intasome, favorisant l'intégration du virus en cellules¹¹⁵. Ceci est en corrélation avec le fait que le VIH-1 s'intègre dans des régions du génome qui sont dynamiques et où l'accessibilité aux nucléosomes est propice pour que l'intasome puisse facilement s'y ancrer. Le FACT a également été identifié comme favorisant l'intégration du Avian Leukosis Virus (ALV) et en influant sur sa sélectivité¹¹⁶.

Cleavage and polyadenylation factor 6 (CPSF6)

CPSF6 (**Figure 22**) joue un rôle primordial dans l'import nucléaire du CPI en interagissant avec la CA, mais est également impliquée dans le choix du site d'intégration^{15,16}. L'intégration dans des cellules déplétées pour CPSF6 est redirigée dans des régions moins actives transcriptionnellement, suggérant que le rôle de CPSF6 est d'éloigner le CPI de régions situées dans l'hétérochromatine¹⁶. Le rôle de CPSF6 dans la sélectivité sera abordé plus en

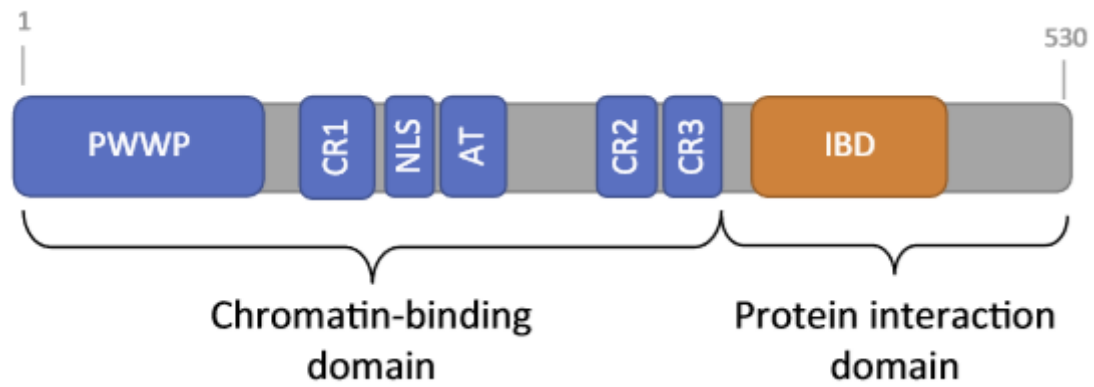


Figure 23 : Organisation schématique de LEDGF/p75. La partie N-terminale de LEDGF (en bleu) est impliquée dans l'interaction avec la chromatine et de modifications d'histones méthylées (domaine PWWP). La partie C-terminale de LEDGF contient le domaine IBD (IN binding domain ; orange) qui permet l'interaction avec l'IN du VIH-1¹¹⁷.

détails dans la partie 4 « la sélectivité d'intégration ».

Lens Epithelium Derived Growth Factor (LEDGF/p75)

LEDGF est le principal cofacteur d'intégration du VIH. C'est une protéine de 530 acides aminés codée par le gène *PSIP1*. LEDGF est un facteur de transcription impliqué dans la recombinaison homologue de l'ADN¹¹⁸. Son rôle est crucial pour la réplication du virus : après l'entrée du CPI dans le noyau, il va le guider dans des régions du génome propices à l'intégration et à l'expression du provirus. Ceci est permis grâce à des domaines d'interaction avec à la fois l'IN du VIH-1 : le domaine IBD¹¹⁹ (integrase binding domain) et la chromatine (le domaine PWWP qui reconnaît certaines modifications d'histones^{120,121}) (**Figure 23**). L'interaction entre LEDGF et l'IN est lentivirus spécifique^{122,123}, son rôle dans le ciblage intervient après celui de CPSF6. Pour résumer brièvement, CPSF6 amène dans un premier temps l'intasome dans des régions transcrites puis LEDGF agit plus localement au niveau du site final d'intégration en guidant l'intasome sur des nucléosomes présents dans les unités de transcription actives où on retrouve entre autre la modification H3K36me3 reconnue par le domaine PWWP de LEDGF. Comme pour CPSF6, le rôle de LEDGF dans la sélectivité sera abordé en détails dans la partie sélectivité.

GCN2

GCN2 est une kinase cellulaire phosphorylant eIF2 α sous certaines conditions de stress cellulaire, régulant ainsi la traduction de manière générale¹²⁴. GCN2 a également été identifiée au laboratoire comme phosphorylant et interagissant avec l'IN du VIH-1 (ainsi que d'autres IN rétroviraux) *in vitro*¹²⁵. L'équipe a également observé que la déplétion de GCN2 dans les cellules conduisait à une augmentation de l'infectivité et que ceci était corrélé à une augmentation de l'intégration de l'ADN viral, suggérant que GCN2 puisse agir comme un facteur de restriction. Bien que pour le moment la raison pour laquelle une IN phosphorylée par GCN2 catalyse moins efficacement l'intégration dans les cellules n'est pas encore connue, le laboratoire propose que ceci soit dû à une interaction diminuée entre l'IN et un cofacteur cellulaire. Malgré le fait que GCN2 semble faire partie des défenses innées contre l'infection au VIH-1 (entre autre), il n'y a pas pour le moment d'antagoniste viral identifié contre GCN2, c'est pourquoi il n'a pas le titre de facteur de restriction actuellement.

RAD51

RAD51 est une protéine impliquée dans la recombinaison homologue de l'ADN lors de cassures double brins. RAD51 a la capacité de se polymériser sur l'ADN et de former un filament nucléoprotéique¹²⁶. Le laboratoire a montré que l'infection de cellules exprimant un haut niveau de RAD51 conduisait à une diminution de la réplication virale et de l'intégration⁹⁷. Dans cette étude l'équipe a pu mettre en évidence une corrélation entre l'activité de RAD51 et sa capacité à inhiber l'intégration en cellules grâce à des modulateurs chimiques.

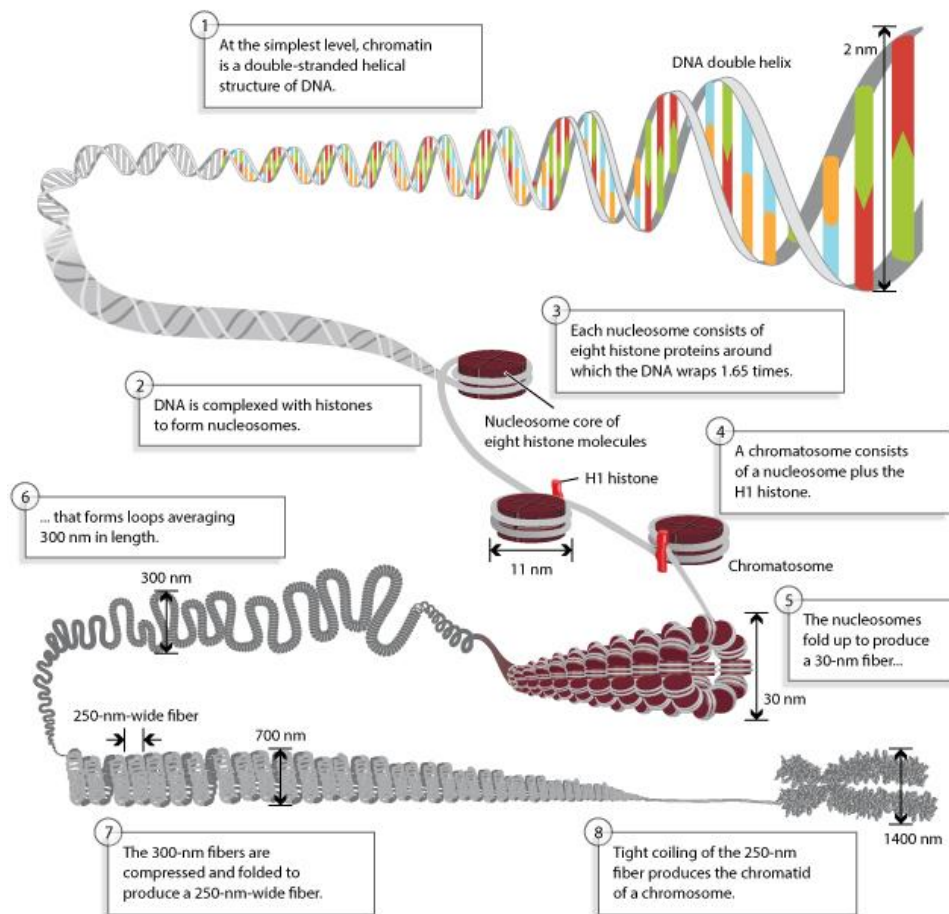


Figure 24 : Représentation des différents niveaux de compaction de l'ADN : de la double hélice d'ADN jusqu'au chromosome. (<https://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310#>).

3 La chromatine

Bien que l'intégration soit dépendante de nombreuses protéines cellulaires et virales comme nous l'avons vu jusqu'à présent, elle est également influencée par la chromatine en elle-même. Certains rétrovirus vont préférer des zones condensées là ou d'autres au contraire vont les éviter. Les protéines qui constituent la chromatine, les histones, jouent aussi un rôle important dans ce ciblage : récemment il a été montré que l'intasome de PFV pouvait interagir directement avec certaines protéines histones, et que ces interactions étaient requises pour une intégration optimale¹²⁷. Plus localement au niveau du substrat final de l'intégration qu'est le nucléosome, la courbure de l'ADN ainsi que la séquence nucléotidique ont également été montrées comme influant sur l'intégration¹²⁷⁻¹²⁹. La structure et les composants de la chromatine jouant un rôle central dans l'intégration rétrovirale, elles seront décrites dans un premier temps puis il sera abordé plus en détails la manière dont cette chromatine peut réguler l'intégration.

3.1 Généralités

a) Composition

La chromatine est composée d'ADN et de protéines histones et non histones servant de support génétique pour la transcription des ARN messagers. Les protéines histones s'assemblent en octamères pour former le nucléosome, qui est le constituant de base de la chromatine, autour duquel 146pb d'ADN sont enroulées¹³⁰. Les nucléosomes sont retrouvés tout au long de l'ADN et permettent de compacter la chromatine, permettant ainsi à un ADN de plus de deux mètres de tenir dans un noyau de quelques micromètres. Différents niveaux de condensation peuvent être retrouvés (**Figure 24**) : le collier de perle (condensation faible induite par la présence des nucléosomes), la fibre de chromatine (niveau de condensation plus élevé grâce à la présence des histones de liaison H1) et le chromosome (niveau de condensation le plus élevé où la fibre de chromatine s'enroule sur elle-même lors de la mitose). Quand la cellule n'est pas en mitose, on distingue deux types de chromatine : l'euchromatine, peu condensée, permet l'expression active des gènes car elle est riche en complexes de remodelages et en facteurs de transcriptions, alors que l'hétérochromatine est plus condensée et l'expression génique est très faible voir nulle.

b) Le nucléosome : substrat final de l'intégration

Le nucléosome est une structure de 146pb d'ADN enroulées autour d'un octamère d'histones dont la structure a été résolue en 1997¹³⁰ (**Figure 25**). Sur cette structure on peut



Figure 25 : Structure cristallographique du nucléosome. Le nucléosome est composé de 146 paires de base d'ADN (en marron et turquoise) enroulées autour d'un octamère d'histones (H2A : jaune ; H2B : rouge ; H3 : bleu ; H4 : vert)¹³⁰.

voir que le cœur du nucléosome est composé d'un octamère d'histones où chacune des histones H2A, H2B, H3 et H4 sont retrouvées en deux exemplaires. Les protéines histones sont riches en lysines et en arginines (résidus basiques) qui sont chargées positivement au pH physiologique, ce qui permet une interaction forte avec les groupements phosphates de l'ADN chargés négativement. Les protéines histones sont composées en C-terminal d'un domaine globulaire qui a une structure permettant la dimérisation des histones entre elles¹³¹. Les extrémités N-terminales sont appelées queues d'histones, car elles dépassent à l'extérieur de l'ADN. Ces queues d'histones ne sont pas structurées et n'apparaissent pas dans la structure du nucléosome car elles sont trop désordonnées. Ces queues d'histones sont la cible de modifications post-traductionnelles qui participent à la régulation de la structure chromatinienne, mais aussi à l'expression des gènes en modulant le recrutement de facteurs de transcription.

3.2 Les modifications post-traductionnelles des queues d'histones

Les modifications post-traductionnelles (MPT) sont des modifications chimiques de protéines ou d'ADN réalisées par des enzymes spécifiques. Ces MPT peuvent avoir différentes conséquences sur la protéine ou l'ADN en fonction de la nature de la modification. Certaines MPT sont même utilisées comme marqueurs pour définir des régions de l'euchromatine ou de l'hétérochromatine. Bien qu'il existe de nombreuses MPT différentes, seules les plus communément retrouvées seront décrites ici pour ce qui est des queues d'histones.

a) La méthylation

L'ajout d'un groupement méthyl est catalysé par les méthyl transférases en utilisant du S-adenosyl méthionine comme substrat donneur de groupement méthyl. Ces méthylations ont lieu principalement sur les résidus arginine et lysine des queues d'histones et plus rarement sur des glutamines et des prolines^{132,133}. Les méthyl transférases contiennent un domaine SET requis pour l'activité catalytique de ces enzymes¹³⁴. A l'inverse, il existe des déméthylases qui peuvent enlever des groupements méthyls, montrant ainsi que la régulation de la méthylation est un processus dynamique dans les cellules¹³⁵. L'ajout de groupements méthyls ne change pas la charge du résidu sur lequel il se fait, c'est une des raisons pour lesquelles la méthylation peut à la fois être une marque épigénétique activatrice ou répressive de la transcription, en fonction de la position et du nombre de groupements méthyls¹³⁶. Certains domaines de protéines appartenant à la « Royal family » sont décrits pour avoir la capacité de reconnaître des histones méthylées comme les chromodomaines, domaines Tudor, MBT, chromo barrel et PWWP¹³⁷.

b) L'acétylation

L'acétylation est catalysée par les acétyl transférases en utilisant l'acétyl coenzyme A comme substrat donneur et se fait sur les lysines¹³⁸. Comme pour les méthylations, l'acétylation est dynamique car il existe des histones déacétylases. De manière générale l'acétylation des histones est associée à la transcription des gènes car cela entraîne une diminution de charge (la lysine chargée positivement devient neutre), et ceci diminue l'interaction entre l'histone et l'ADN, aboutissant à une structure moins compacte et une transcription des gènes plus importante¹³⁹. Parmi les protéines qui interagissent avec les histones acétylées sont retrouvées les protéines contenant un bromodomaine telles que les protéines BET.

c) La phosphorylation

La phosphorylation et la déphosphorylation sont catalysées par les kinases et les phosphatases respectivement. L'addition de groupements phosphates se fait en utilisant l'ATP comme substrat, ce qui introduit une charge négative sur le résidu phosphorylé¹⁴⁰. Elle a lieu principalement sur les sérine, thréonine et tyrosine. L'addition d'une charge négative sur les histones entraîne une relaxation de la chromatine, ce qui permet une meilleure transcription des gènes. La phosphorylation est également impliquée dans la réponse des dommages à l'ADN, car en augmentant l'accessibilité à l'ADN, les facteurs de réparation ont plus de facilités à atteindre le site endommagé^{141,142}.

d) L'ubiquitination

L'ubiquitination est catalysée par une ubiquitine ligase sur les résidus lysines des queues H2A et H2B en grande majorité. Même s'il existe plusieurs formes d'ubiquitination, celle retrouvée sur les histones est en très grande partie de la mono-ubiquitination et est aussi bien associée à l'activation qu'à la répression de la transcription en fonction de la position du résidu qui est modifié¹⁴³.

3.3 Dynamique de la chromatine

Les processus cellulaires associés à l'ADN tels que la réplication, la transcription des gènes ou la réparation sont grandement limités au sein d'une chromatine compactée. Pour pallier à ce problème, des complexes de remodelage sont capables d'altérer l'interaction entre les histones et l'ADN de manière transitoire et locale afin d'assurer le bon déroulement de ces

processus et au bon moment^{144,145}. Ces zones dynamiques de la chromatine sont favorables pour l'intégration de certains rétrovirus tel que le VIH, il n'est donc pas surprenant que certains de ces complexes de remodelage aient été identifiés comme importants pour l'intégration de ces rétrovirus comme les complexes SWI/SNF¹⁰⁷ et FACT^{115,116}.

Deux mécanismes d'actions principaux sont proposés pour les facteurs de remodelage. Le premier est celui de glissement ou « sliding » où l'octamère d'histone est déplacé le long de l'ADN sans en être dissocié, même si de nombreux contacts histone/ADN doivent alors être rompus puis reformés de manière coordonnée¹⁴⁶ (c'est le cas du complexe SWI/SNF). L'autre mécanisme est plus drastique pour le nucléosome car il consiste en l'éviction de certaines histones de l'octamère. L'éviction peut être totale (l'octamère entier) ou partielle (uniquement certaines histones). L'ADN n'étant plus en contact avec les protéines histones, il devient alors beaucoup plus accessible (c'est le cas du complexe FACT, qui est capable de dissocier un dimère H2A/H2B de l'octamère puis de le réassocier)¹⁴⁷.

Retroviridae genera	in Transcription Units ^a	± 2 kb Transcription Start Sites ^a	± 2 kb CpG Islands ^a
Lentiviruses	+/++	0	-/0
Alpharetroviruses	+	0	+
Betaretroviruses	0	0	0
Gammaretroviruses	+	++	++
Deltaretroviruses	+	+	+
Epsilonretroviruses	NA	NA	NA
Spumaviruses	0	++	++

Figure 26 : Tableau résumant les préférences des différents rétrovirus par rapport à certaines régions de la chromatine. Les préférences pour les différentes régions de la chromatine sont représentées par 0 (pas de différence significative par rapport au random), +/++ (favorisée significativement), - (défavorisée significativement) et NA (non disponible)¹⁴⁸.

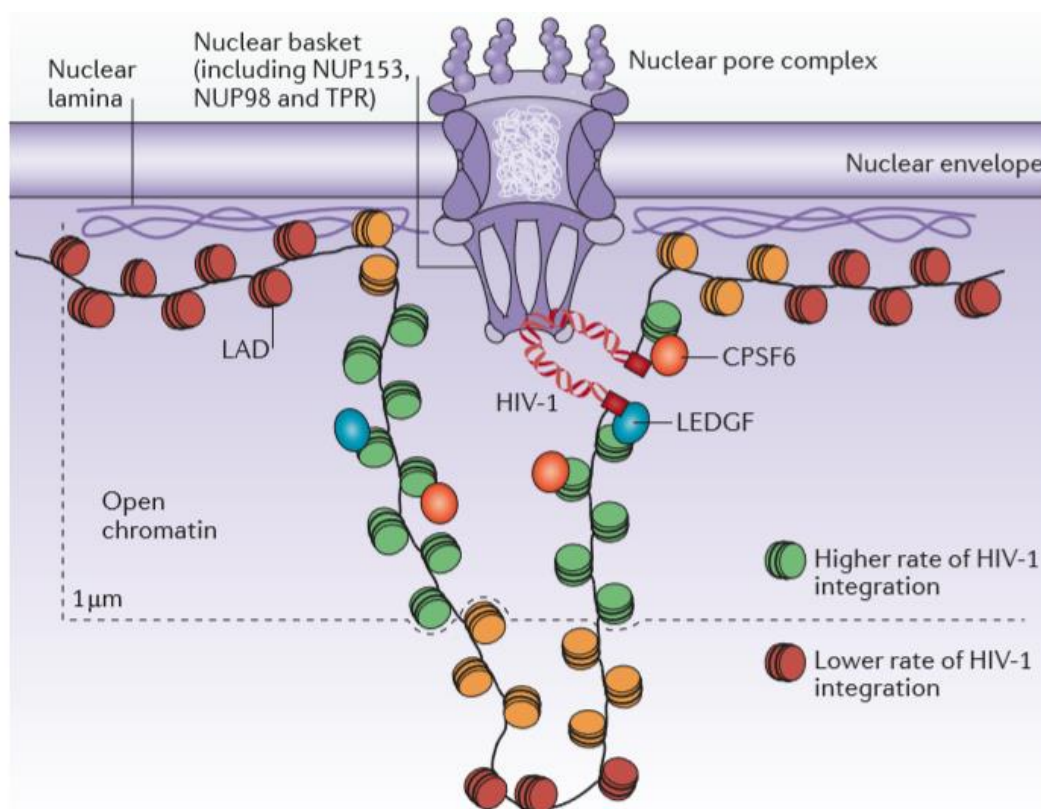


Figure 27 : Intégration du VIH-1 à proximité de la périphérie nucléaire. La fréquence d'intégration du VIH-1 en fonction de la proximité du pore nucléaire est représentée par un code couleur (vert : très fréquent ; orange : fréquent ; rouge : peu fréquent)⁴¹.

4 La sélectivité d'intégration

L'analyse de la distribution des sites d'intégration dans les cellules infectées a permis de mettre en évidence que l'intégration ne se faisait pas de manière aléatoire dans la chromatine et que ces préférences sont spécifiques pour chaque genre de rétrovirus (**Figure 26**). Les mécanismes qui influent sur le choix du site d'intégration constituent la sélectivité d'intégration. Les lentivirus s'intègrent préférentiellement dans les unités de transcription active¹⁴⁹ à l'inverse des spumavirus qui préfèrent les domaines associés à la lamine¹²⁷ qui sont des régions condensées de la chromatine¹⁵⁰. Les gammarétrovirus ont une préférence pour les régions promotrices. Et enfin pour les alpharétrovirus et les bétarétrovirus il ne semble pas qu'il y ait de préférence pour des régions particulières^{148,151}. Même si ces différences de ciblage ne sont pas encore totalement expliquées, plusieurs facteurs sont proposés comme étant impliqués dans cette sélectivité : la manière dont le CPI est importé dans le noyau, l'accessibilité à la chromatine et des interactions du CPI avec des facteurs de ciblage.

4.1 L'import nucléaire

Les lentivirus ont la capacité d'infecter des cellules qui ne se divisent pas grâce à l'import actif du CPI dans le noyau grâce majoritairement à l'interaction de la CA virale avec certaines protéines cellulaires (NUP153, NUP358, CYPA et CPSF6) dont chacune est impliquée dans l'import nucléaire du CPI et dans le choix du site d'intégration¹⁰⁹. Dans des cellules déplétées pour NUP358, l'intégration est diminuée dans les zones condensées de la chromatine¹⁵². De même pour un mutant de la CA n'interagissant plus avec CPSF6^{50,153} (N74D ou A77V). En revanche l'effet opposé a été observé dans le cas de la perte d'interaction entre la CA et CYPA où l'intégration était redirigée dans des zones condensées de la chromatine¹⁹.

Là où la déplétion de LEDGF entraîne une diminution de l'intégration dans les gènes de 83% (dans les cellules normales) à environ 63% (dans les cellules KO LEDGF), la déplétion de CPSF6 entraîne une baisse de ce pourcentage à 57%. Les virus s'intégrant dans des cellules déplétées pour les deux facteurs à la fois ne retiennent que 48% d'intégration dans les gènes, ce qui est très proche de la valeur d'une distribution aléatoire (qui est de 44,7%). Ces données soulignent à quel point CPSF6 et LEDGF sont importants pour la sélectivité du VIH-1¹⁶.

4.2 Choix du site final

Le modèle actuel propose que le CPI interagisse dans un premier temps avec CPSF6 au niveau des pores nucléaires pour faciliter son import dans le noyau et l'amener vers des régions chromatiniennes peu denses. L'intégration du VIH se fait majoritairement dans la chromatine à proximité de la périphérie nucléaire^{154,155} (**Figure 27**) où NUP153^{48,156} et LEDGF^{157,158} ont été montrés comme requis pour ce ciblage.

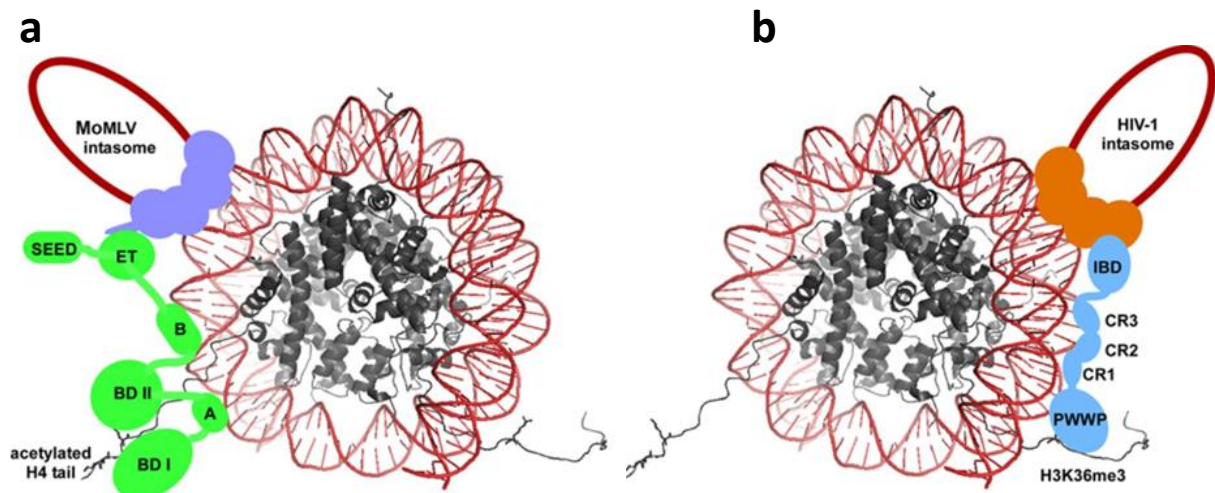


Figure 28 : Modèle montrant l'interaction bimodale des protéines BET de et LEDGF avec leur intasome correspondant MLV et VIH-1 sur un mononucleosome. (a) Les protéines BET (en vert) sont capables d'interagir avec l'ADN nucléosomal et des d'histones acétylées. Leur région C-terminale est également capable d'interagir avec l'intasome de MLV. **(b)** LEDGF (en bleu) interagit avec l'ADN nucléosomal ainsi que des histones méthylées issues de la queue H3. Le domaine IBD en C-terminal lui interagit avec l'intasome du VIH-1. (Les protéines histones sont en gris avec l'ADN nucléosomal enroulé autour en rouge)¹⁵⁹.

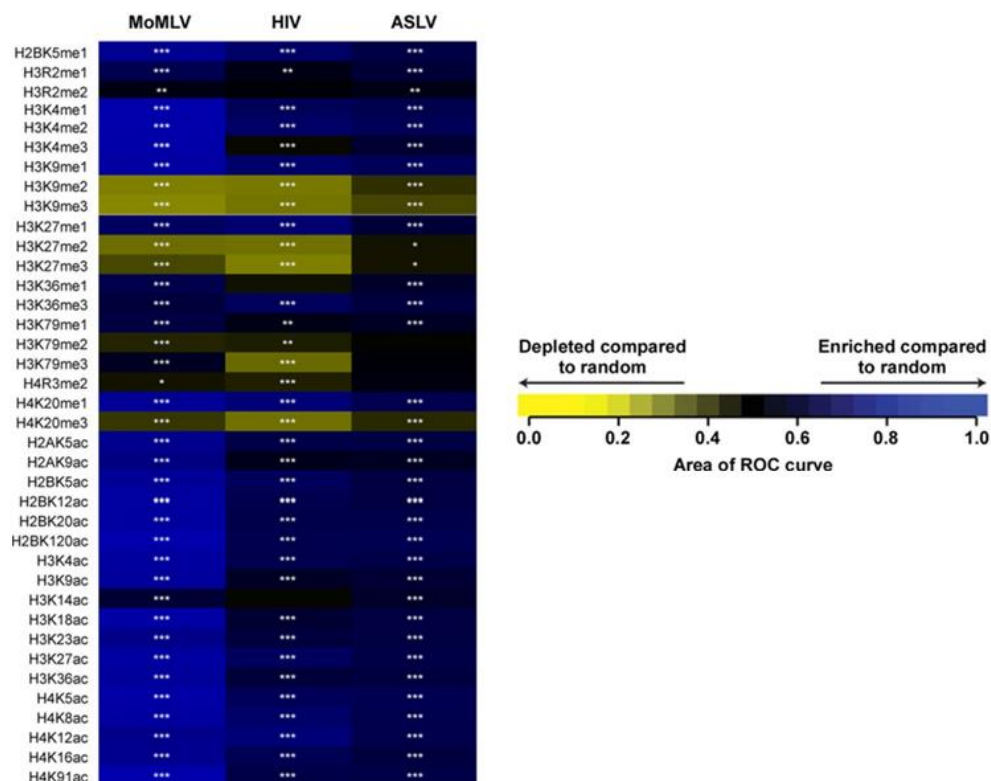


Figure 29 : Heatmap indiquant la relation entre la fréquence d'intégration de MLV, VIH et ASV et la nature des modifications post traductionnelles des histones. La fréquence d'intégration d'un rétrovirus pour une modification donnée est représentée par un code couleur (à droite de l'image). Les étoiles représentent la significativité par rapport à un contrôle random avec *** $P < 0,001$; ** $P < 0,01$; * $P < 0,05$ ¹⁵⁹.

Le CPI est ensuite guidé sur des nucléosomes (substrat favorisé comparé à de l'ADN non chromatinisé) qui présentent des modifications d'histones associées à la transcription active des gènes telles que H3K36me3 et H4K20me1¹⁵⁹. Il est supposé que ce ciblage final sur le nucléosome est principalement dû à l'interaction du CPI avec certains cofacteurs cellulaires qui à leur tour reconnaissent certaines modifications épigénétiques. De tels cofacteurs ont été identifiés et assez bien décrits pour les gammarétrovirus et les lentivirus.

Les gammarétrovirus s'intègrent préférentiellement dans les zones promotrices qui sont caractérisées entre autre par la présence de modifications d'histones acétylées¹⁶⁰⁻¹⁶². Les protéines BET (bromo and extra terminal domain) telles que BRD2, 3 et 4 ont été identifiées comme jouant un rôle central dans ce ciblage¹⁶³⁻¹⁶⁵. Différentes troncations des protéines BET et de l'IN de MLV ont permis d'identifier des zones d'interaction entre ces deux protéines : le domaine ET des protéines BET est apparu comme étant un déterminant majeur, même si le domaine BID participe également, il n'est pas requis mais permet une liaison plus forte¹⁶⁶. Du côté de l'IN, c'est la partie C-terminale qui est majoritairement impliquée. Les IN des gammarétrovirus possèdent une caractéristique qui leur est propre : une queue C-terminale de 28 acides aminés dont la fonctionnalité a longtemps été un mystère¹⁶⁷. Cette région est unique car elle sert pour l'interaction avec le facteur de ciblage des gammarétrovirus^{165,166}. Après interaction de l'intasome avec les protéines BET, ces dernières vont diriger l'intasome dans des régions promotrices par reconnaissance de modifications épigénétiques présentes dans ces régions. Ceci est médié par les bromodomaines I et II (**figure 28a**) qui reconnaissent spécifiquement des lysines acétylées présentes sur les queues d'histones H3 et H4¹⁶⁸⁻¹⁷⁰. La déplétion des protéines BET ou l'utilisation de molécules empêchant l'interaction des BET avec les modifications d'histones qu'elles reconnaissent, entraînent dans les deux cas un changement de sélectivité avec une diminution de la proportion des sites d'intégration dans les régions promotrices^{163,165}.

Un mécanisme très proche est utilisé chez les lentivirus avec une interaction bimodale du cofacteur LEDGF avec à la fois l'intasome lentiviral et la chromatine (**Figure 28b**). Les lentivirus s'intègrent dans les unités de transcription de la chromatine qui sont enrichies en modifications d'histones associées à la transcription des gènes (**Figure 29**). Le domaine IBD de LEDGF interagit avec l'IN au niveau de l'interface de dimérisation entre deux CCD de l'IN¹⁷¹, et le domaine PWWP reconnaît majoritairement la modification d'histone H3K36me3^{120,121}. Il existe une isoforme de LEDGF (LEDGF/p52) mais qui ne contient pas de domaine IBD et par conséquent ne peut pas interagir avec l'IN¹⁷². En revanche il existe une autre protéine cellulaire contenant à la fois un domaine IBD et PWWP : HRP2 (**Figure 30**), appartenant à la même famille que LEDGF/p75. Les premières analyses des sites d'intégration dans des cellules déplétées pour LEDGF/p75 ont été assez contradictoires car là où certains laboratoires voyaient un effet conséquent, d'autres n'observaient pas d'effet significatif sur la présence ou non de LEDGF. Ces différences de résultats peuvent aujourd'hui être expliquées de plusieurs manières : il est estimé qu'il y a environ 100 000 molécules de LEDGF par cellule¹⁷³, ce qui est de beaucoup supérieur à ce qu'il faut pour l'intégration des CPI. Par conséquent des traces résiduelles de LEDGF peuvent être suffisantes pour supporter un profil d'intégration proche

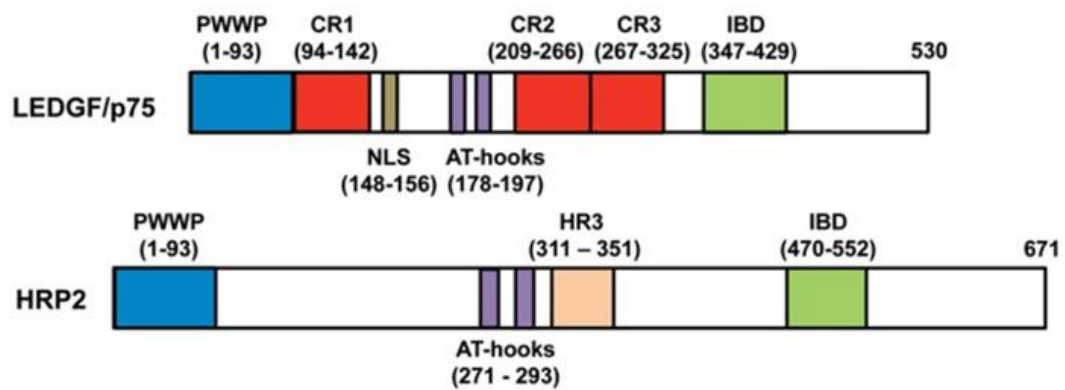


Figure 30 : Organisation schématique de LEDGF/p75 et de HRP2. LEDGF/p75 et HRP2 appartiennent toutes les deux à la même famille et possèdent également toutes les deux un domaine PWWP (en bleu) ainsi qu'un IBD (en vert) fonctionnels¹⁵⁹.

des cellules WT. L'autre explication est que HRP2 est capable d'assurer en partie la fonction de LEDGF quand celui-ci n'est pas présent. En effet la double déplétion de LEDGF et HRP2 aboutit à une plus grande diminution des sites d'intégration dans les unités de transcription (même si dans ce cas l'intégration se fait toujours préférentiellement dans les unités de transcription) que la seule déplétion de LEDGF^{174,175}. Ces résultats montrent à quel point la sélectivité est un processus complexe et que de nombreux facteurs régulent cette sélectivité.

4.3 Impact de la structure de l'ADN nucléosomal

L'ADN enroulé autour de l'octamère d'histone présente des caractéristiques différentes en fonction de sa position sur le nucléosome et de la nature des nucléotides qui le composent. En plus de l'alternance entre petits et grands sillons, il existe des motifs nucléotidiques qui sont plus ou moins flexibles, et cette flexibilité est importante pour l'intégration. Ainsi, le choix du site d'intégration final ne se fera pas non plus de manière aléatoire sur un nucléosome donné. Le motif pyrimidine-purine (YR) et RY sont respectivement les motifs les plus et les moins flexibles alors que les motifs RR et YY présentent des niveaux intermédiaires de flexibilité. Il semble qu'il y ait une corrélation entre la flexibilité du motif et la duplication de paire de base générée par l'intégration¹⁷⁶ : il a été montré que les virus générant une duplication de paire de base de 4 s'intègrent préférentiellement au niveau de motifs très flexibles YR car la courbure naturelle de l'ADN nucléosomal n'est pas propice à l'intégration de ces rétrovirus. En effet, le complexe de transfert de brin de PFV est capable de tordre cet ADN nucléosomal afin de lui donner une structure compatible pour l'intégration¹²⁷. Les virus générant une duplication de paire de base de 5 sont plus intermédiaires avec tout de même une préférence pour les motifs YR ou YY. Et pour finir les virus générant une duplication de paire de base de 6 semblent préférer les extrêmes RY et YR. Ces données suggèrent que ce n'est pas forcément la séquence primaire en nucléotides qui importe, mais plutôt des combinaisons de bases qui présentent une plus ou moins grande flexibilité.

4.4 Impact de la condensation de la chromatine

Les cofacteurs cellulaires ne sont pas les seuls à jouer un rôle dans la sélectivité, l'IN elle-même semble impliquée. En effet, *in vitro* et en absence de cofacteurs ou de modifications épigénétiques, notre laboratoire a pu montrer que les IN de différents rétrovirus avaient des comportements différents vis-à-vis de la compaction de la chromatine¹⁷⁷. Alors que PFV et MLV préfèrent des structures nucléosomales denses et compactes, VIH et ASV s'intègrent mieux dans des structures de faible densité en nucléosomes *in vitro*. Au vu de ces différences de comportement, il est supposé que l'architecture de l'intasome et la compaction de la chromatine sont des paramètres importants pour la sélectivité. Ceci a été validé par un

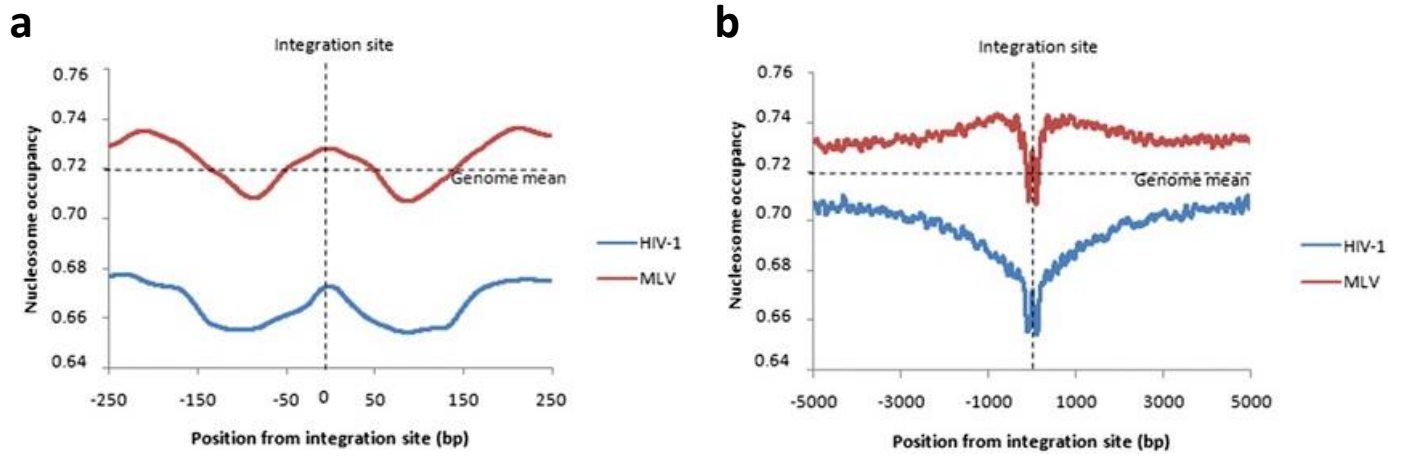


Figure 31 : Corrélation entre la position des sites d'intégration du VIH-1 et de MLV et la densité en nucléosomes. (a) Moyenne de densité en nucléosomes encadrant le site d'intégration pour le VIH-1 ou MLV (entre -250 paires de base et +250 paires de base). **(b)** Moyenne de densité en nucléosomes encadrant le site d'intégration pour le VIH-1 ou MLV (entre -5000 paires de base et +5000 paires de base)¹⁷⁷.

algorithme permettant de prédire la position des nucléosomes par rapport aux sites d'intégration de différents rétrovirus (**Figure 31**). Dans cette analyse, il a été montré par exemple que MLV s'intègre dans des régions de plus haute densité nucléosomale que VIH dans les cellules, mais que pour les deux rétrovirus l'environnement proche du site d'intégration est de plus faible densité bien qu'un nucléosome reste présent sur le site précis d'intégration. Ces données démontrent que l'environnement proche du nucléosome ciblé par le virus possède des caractéristiques bien précises et est un paramètre important pour le choix du site d'insertion. Il est également supposé que dans le cas du VIH et de ASV, la condensation de la chromatine puisse masquer des contacts entre l'intasome et le nucléosome requis pour une intégration efficace. Ces données corroborent avec le fait que des complexes de remodelage de la chromatine tels que SWI/SNF¹⁰⁷ et le FACT^{115,116} favorisent l'intégration du VIH (et de ASV pour ce qui est du FACT).

5 Objectifs de l'étude

Ces dernières années les travaux du laboratoire se sont focalisés sur l'intégration rétrovirale, afin de mieux comprendre ce processus et les mécanismes qui la régit. Nous avons pu le voir jusqu'à présent, l'intégration des rétrovirus dépend de nombreux paramètres dont beaucoup ne sont pas encore complètement identifiés et compris. Les données de la littérature et issues du laboratoire montrent que l'interaction entre l'intasome et la chromatine est un élément clé de l'intégration. La régulation de ce processus par la structure chromatinienne et son remodelage suggère que des contacts directs protéine/protéine entre l'intasome et le nucléosome sont impliqués.

Ceci est renforcé par la découverte au laboratoire d'une interaction directe entre l'intégrase du VIH-1 et une queue d'histone (la queue H4) *in vitro*¹⁷⁸. Etant donné que les queues d'histones sont retrouvées sur le substrat final de l'intégration qu'est le nucléosome, le laboratoire s'est demandé si cette interaction pouvait jouer un rôle dans l'intégration du VIH-1. Les queues d'histones sont déjà connues pour être impliquées dans le choix du site d'intégration de plusieurs rétrovirus en interagissant avec des cofacteurs d'intégration, mais n'ont à ce moment jamais été identifiées comme participant directement à des interactions avec les intasomes.

Le projet de cette thèse s'est donc focalisé autour de cette interaction intégrase-queue d'histone H4, afin de déterminer dans un premier temps si elle participe au cycle de réplication du virus et si elle constitue une interaction hôte-pathogène fonctionnelle. Dans un deuxième temps, nous avons analysé en détails les mécanismes moléculaires qui régissent cette interaction, afin de comprendre ses implications lors de l'intégration du VIH. Enfin, dans un dernier temps nous nous sommes focalisés sur le développement d'une approche pharmacologique visant à empêcher cette interaction intégrase-queue d'histone H4 avec des composés.

Ce projet a pour but de mieux comprendre le fonctionnement de l'intégration rétrovirale en identifiant de nouveaux partenaires potentiels de l'intasome du VIH-1, et de comprendre comment ces partenaires sont capables d'influer sur l'intégration. Si cette interaction se trouve être importante pour le bon déroulement du cycle viral, elle constituerait ainsi une nouvelle interaction hôte-pathogène qui pourrait être utilisée comme base pour le développement de nouvelles stratégies antivirales.

MATERIELS ET METHODES

1 Matériels

1.1 Culture cellulaire et bactérienne

a) Les cellules

- **HEK 293T**

Les cellules HEK 293T sont une lignée de cellules épithéliales semi-adhérentes issues de cellules embryonnaires de rein humain possédant l'antigène T du virus SV40. Elles sont facilement transfectables et sont utilisées pour la production des vecteurs lentiviraux ainsi que pour les études des phases précoces de la réplication virale.

- **HeLa P4**

Les cellules HeLa P4 sont une lignée de cellules cancéreuses adhérentes exprimant de manière constitutive le récepteur CD4 à leur surface. Elles sont transformées de manière stable avec le gène LacZ sous contrôle du promoteur viral LTR.

- **K562**

Les cellules K562 sont une lignée de cellules cancéreuses indifférenciées qui poussent en suspension. Les caractéristiques génomiques (modifications d'histones, position des nucléosomes...) de cette lignée cellulaire sont particulièrement bien annotées, et nous ont servi pour les analyses de sélectivité.

- **TZM et TZM KO LEDGF**

Les cellules TZM sont dérivées d'un clone de cellule HeLa qui a été modifié pour exprimer diverses protéines de surface (CD4, CCR5 et CXCR4) et contient également un gène rapporteur pour la luciférase et la bêta galactosidase, sous contrôle du LTR du VIH-1. Cette lignée a été utilisée pour réaliser une lignée KO LEDGF par TALEN (Transcription Activator-like Effector Nuclease), afin d'étudier la réplication virale dans un contexte indépendant de LEDGF.

b) Les bactéries

- **DH5α**

Les bactéries E.coli DH5α sont utilisées pour l'amplification des plasmides car elles sont très facilement transformables. Une mutation des gènes endA1 et recA1 permet l'inactivation d'endonucléases intracellulaires qui peuvent dégrader l'ADN plasmidique dans certaines méthodes de miniprep.

- **BL21 (DE3)**

Les bactéries BL21 (DE3) PlysS sont utilisées pour la production des protéines telles que les IN rétrovirales. Elles contiennent un plasmide PlysS codant pour l'ARN polymérase T7 sous contrôle du promoteur lac UV5 qui permet l'induction de l'expression de la protéine d'intérêt.

1.2 Plasmides

a) Plasmide d'expression de l'IN du VIH-1

Le plasmide utilisé pour la purification de l'IN du VIH est le pET-21b-IN. La région du gène *pol* de l'IN du VIH-1 est sous contrôle du promoteur de l'ARN polymérase T7 qui peut être sélectionné grâce à l'ampicilline.

b) Plasmides pour la production des vecteurs lentiviraux

La production des vecteurs lentiviraux pseudotypés VSVG est assurée par la plateforme de vectorologie de Bordeaux et requiert la transfection de 3 plasmides différents :

- **Le plasmide de transfert**

Ce plasmide permet la transcription d'un ARN mimant l'ARN génomique viral contenant les éléments permettant la production des particules virales (une séquence d'encapsidation psi) ainsi que les phases précoces de la réplication (une région cPPT pour la transcription inverse et les LTR pour l'intégration). Ces vecteurs effectuent les étapes du cycle viral jusqu'à l'intégration : ils sont non réplcatifs grâce à une délétion dans la région U3 des LTR. Ce plasmide contient également un gène codant pour l'eGFP sous contrôle du promoteur PGK.

- **Le plasmide d'empaquetage**

Ce plasmide contient les séquences codantes pour les gènes *gag* et *pol* du VIH ainsi que pour les protéines accessoires Tat et Rev. Ces gènes sont sous dépendance d'un promoteur CMV. La production des vecteurs contenant une IN mutée se fait à partir de ce plasmide en mutant la séquence codante pour l'IN.

- **Le plasmide d'enveloppe**

Ce plasmide code pour la protéine G du VSV qui va remplacer l'enveloppe naturelle du VIH afin d'augmenter le tropisme des vecteurs lentiviraux, nous permettant ainsi d'infecter un plus large choix de cellules.

1.3 Oligonucléotides

Utilisation	Nom	Séquence (5' → 3')
Génération donneur HIV	HIV- SupF 3'	ACTGGAAGGGGCTAATTCACCTA ACGTTGCCCGGATCCGGTCGCGC
Génération donneur HIV	HIV- SupF 5'	ACTGCTAGAGATTTTCCACAGCG GCGCGTCATTTGATATGATGCG
Génération donneur PFV	PFV- SupF 3'	AATGTAGTCTTATGCAATACTCTTGTAGTCT TGCAAGCGGCGCGTCATTTGATATGATGCG
Génération donneur PFV	PFV-SupF 5'	ATTGTCATGGAATTTTGTATATTGATTAT CCTTTAACGTTGCCCGGATCCGGTCGCGC
Génération donneur RAV	RAV- SupF 3'	AATGTAGTCTTATGCAATACTCT AATGCAATATTAACGTTGCCCG
Génération donneur RAV	RAV- SupF 5'	AATGAAGCCTTCTGCTTCATGCA TGTGCTCGTGCAGCGCGTCATT
Génération donneur MVV	MVV- SupF 3'	GCTGCGAGATCCGCTCCGGTGTGGAC GGTTAACGTTGCCCGGATCCGGTCGCG
Génération donneur MVV	MVV- SupF 5'	GCTGCGAGATCCGCTCCGGTGTGGAC GGGCGGCGCGTCATTTGATATGATGCG
Donneur court HIV	HIV U5 (+)	GTGGAAAATCTCTAGCAGT
Donneur court HIV	HIV U5 (-)	ACTGCTAGAGATTTTCCAC
Test processing	70	GTGTGGAAAATCTCTAGCAGT
Test transfert de brins	71	GTGTGGAAAATCTCTAGCA
Test processing + transfert de brins	72	ACTGCTAGAGATTTTCCACAC
Assemblage intasome MVV	MVV INTASOME U3	CCGTGCA ACACCGGAGCGGATCTCGCA
Assemblage intasome MVV	MVV INTASOME U5	GCTGCGAGATCCGCTCCGGTGT TGCACGG
Assemblage intasome PFV	PFV INTASOME U3	TGCGAAATTCATGACA
Assemblage intasome PFV	PFV INTASOME U5	ATTGTCATGGAATTTTCGCA

1.4 Peptides

Nom	Fournisseur	Séquence
H4K20me0-biot	Eurogentec	S-G-R-G-K-G-G-K-G-L-G-K-G-G-A-K-R- H-R-K-V-L-R-G-G-K(biotin)
H4K20me1-biot	Eurogentec	S-G-R-G-K-G-G-K-G-L-G-K-G-G-A-K-R- H-R-K(me1)-V-L-R-G-G-K(biotin)
H4K20me2-biot	Eurogentec	S-G-R-G-K-G-G-K-G-L-G-K-G-G-A-K-R- H-R-K(me2)-V-L-R-G-G-K(biotin)
H4K20me3-biot	Eurogentec	S-G-R-G-K-G-G-K-G-L-G-K-G-G-A-K-R- H-R-K(me3)-V-L-R-G-G-K(biotin)
H2A-biot	Eurogentec	R-S-S-R-T-K-A-K-A-R-A-K-G-G-Q-K-G-R- G-S-G-G-K(biotin)
H2BK5me1-biot	Eurogentec	P-E-P-A-K(me1)-S-A-P-A-P-K-K-G-S-K- K-A-V-T-K-A-Q-G-G-K(biotin)
H3K36me3-biot	Eurogentec	A-T-K-A-A-R-K-S-A-P-A-T-G-G-V- K(me3)-K-P-H-R-Y-R-P-G-G-K(biotin)
Scramble H4-biot	IGBMC, Strasbourg	K-G-R-S-G-K-L-A-K-G-G-H-G-G-V-K-G- R-G-K(me1)-G-R-R-L-G-K(biotin)
H4K20me0-FITC	IGBMC, Strasbourg	FITC-K-G-G-R-H-R-K-V-L-R
H4K20me1-FITC	IGBMC, Strasbourg	FITC-K-G-G-R-H-R-K(me1)-V-L-R
H4K20me2-FITC	IGBMC, Strasbourg	FITC-K-G-G-R-H-R-K(me2)-V-L-R

H4K20me3-FITC	IGBMC, Strasbourg	FITC-K-G-G-R-H-R-K(me3)-V-L-R
H4K20me1	Eurogentec	S-G-R-G-K-G-G-K-G-L-G-K-G-A-K-R-H-R-K(me1)-V-L-R
H2A	Eurogentec	R-S-S-R-T-K-A-K-A-R-A-K-G-G-Q-K-G-R-G-S
H2BK5me1	Eurogentec	P-E-P-A-K(me1)-S-A-P-A-P-K-K-G-S-K-K-A-V-T-K-A-Q
H3K36me3	Eurogentec	A-T-K-A-A-R-K-S-A-P-A-T-G-G-V-K(me3)-K-P-H-R-Y-R-P

1.5 Anticorps

Nom	Clone	Cible	Fournisseur	Espèce	Dilution
IN	Polyclonal	IN du VIH-1	BioProducts	Lapin	1/5000
HIS	Monoclonal	Tag histidine	Novagen	Souris	1/1000
GST	Monoclonal	Tag GST	ThermoFisher	Lapin	1/5000

1.6 Nucléosomes

a) Nucléosomes commerciaux

Les mononucléosomes utilisés pour les tests d'intégration et les pull down sont achetés chez EPICYPHER. Ils sont assemblés avec des histones humaines recombinantes exprimées dans *E. coli* et un ADN de 147pb (la séquence 601) dont l'extrémité 5' est biotinylée.

b) Nucléosomes « tail less »

Les mononucléosomes délétés pour une des queues d'histone (tail less) ont été assemblés au laboratoire. Pour cela, 10µg d'octamères pour lesquels il manque une queue d'histone (qui nous ont été fournis par le laboratoire de K. Lüger) sont incubés avec 5µg d'ADN 601 biotinylé dans des conditions de forte salinité (2M NaCl ; Tris pH 7,7 ; 0,5M EDTA) et dialysés sur la nuit. Cette première dialyse sera suivie par 3 dialyses successives de 3h dans des solutions dont la salinité décroît (0,85M ; 0,65M puis 0,2M NaCl). Une dernière dialyse est effectuée sur la nuit dans un tampon qui ne contient pas de sel (Tris pH 7,7 ; 0.5M EDTA). L'efficacité des assemblages est contrôlée par migration des produits sur gel natif PAGE 8% et révélation au SYBR Safe.

1.7 Composés chimiques

Les molécules identifiées par le crible alphascreen, l'auranofine et la nisoldipine, ont été achetées chez Sigma-Aldrich sous forme de poudre, et solubilisées dans du DMSO.

La molécule ayant eu le meilleur score lors du criblage virtuel (nom chimique : 2-naphthylBLAH) a été achetée chez Sigma-Aldrich et solubilisée dans du DMSO.

1.8 Virus

a) Vecteurs lentiviraux pseudotypés VSVG

Les vecteurs lentiviraux pseudotypés VSVG sont des virus recombinants intégratifs mais non réplicatifs, on peut ainsi les utiliser en routine dans un laboratoire de type P2 pour l'étude des phases précoces de la réplication virale. De plus, ces vecteurs nous permettent d'introduire des mutations dans la séquence codante pour l'IN et d'étudier l'impact de ces mutations sur l'infectivité de ces virus mutants. Ces vecteurs possèdent un gène rapporteur, l'eGFP, qui exprimeront cette protéine si le génome viral est intégré dans le génome de l'hôte car l'eGFP est sous dépendant d'un promoteur interne PGK. Nous pouvons ainsi suivre l'infectivité en mesurant le pourcentage de cellules GFP+ par cytométrie en flux.

b) Le VIH-1 sauvage

La souche de virus sauvage utilisée dans cette étude nous a été fournie par le CHU de Bordeaux, ce virus est issu de la co-culture de PBMC naïves avec des PBMC d'un patient diagnostiqué VIH+. Le surnagent de cette co-culture a ensuite été récupéré après 3 semaines et stocké à -80°C.

2 Méthodes

2.1 Culture bactérienne

a) Culture

Les pré-cultures bactériennes liquides sont réalisées sous agitation (220 rpm) à 37°C dans du milieu LB sur la nuit. En fonction du gène de résistance, la sélection est faite avec de l'ampicilline (100 µg/mL) ou de la kanamycine (50 µg/mL). La température de culture pour la purification des protéines est ensuite ajustée en fonction de la protéine.

b) Transformation

Les transformations sont réalisées avec des bactéries chimio-compétentes. Le plasmide à transférer (une centaine de ng) est incubé avec 50 µL de bactéries compétentes dans la glace pendant 5 minutes, puis un choc thermique à 42°C pendant 30 secondes permet l'entrée du vecteur dans les bactéries. Les bactéries transformées sont ensuite mises en culture dans 500 µL de LB à 37°C pendant 1 heure. Finalement, 50µL de cette culture sont étalés sur des boîtes de LB contenant l'antibiotique approprié et sont mises à 37°C sur la nuit.

2.2 Culture cellulaire

a) Culture

Les cellules HEK 293T, HeLa P4, K562 et TZM sont cultivées dans du milieu Dulbecco's modified minimal essential medium (Invitrogen) complémenté avec 10% de sérum de veau fœtal (Invitrogen) et 50 mg/mL de gentamycine (Invitrogen) à 37°C dans une atmosphère à 5% de CO₂. Pour les HeLa P4, 1 mg/mL de G418 est ajouté au milieu pour maintenir l'expression du récepteur CD4. Quand les cellules arrivent à confluence, elles sont décollées avec une solution de trypsine-EDTA (Invitrogen) pour les cellules adhérentes, et sont diluées au 1/10 dans du milieu de culture.

Les PBMC sont une lignée primaire de cellules naturellement infectables par le VIH. Elles sont récupérées à partir d'échantillons frais de sang de patients séronégatifs qui nous sont procurés par l'établissement français du sang (EFS) de Bordeaux. Après isolation des PBMC, les cellules sont cultivées dans du milieu RPMI with glutamax (Gibco) complémenté avec 10% de sérum et veau fœtal décomplémenté, 10 000 unités d'interleukine II, 35 µL d'anti

interféron alpha et 1mL d'un mélange pénicilline/streptomycine. Deux jours avant l'infection, les cellules sont activées avec de la PHA à 5 µg/mL et cultivées à 37°C et 5% de CO₂.

b) Tests de toxicité

La toxicité des composés que l'on veut étudier en cellules est réalisée grâce au kit CellTiter 96 aqueous one solution cell proliferation de Promega. Ce test est effectué en plaques 96 puits où l'on a distribué 200 000 cellules par puits dans un volume final de 200 µL. Une gamme de concentration des composés à tester est ajoutée aux cellules qui sont laissées 72h à 37°C, 5% de CO₂. L'analyse de la toxicité est réalisée par l'ajout du réactif CellTiter (40 µL) et après incubation pendant 2 à 4h à 37°C, l'absorbance de chaque puits est lue à 490nm.

c) Infection par les vecteurs lentiviraux

- **Production**

La production des vecteurs lentiviraux est assurée par la plateforme de vectorologie de Bordeaux. Des cellules HEK 293T (environ 6.10^6 cellules) sont mises en culture dans 10 mL de DMEM pendant 24h à 37°C. Les trois plasmides décrits ci-dessus sont ensuite transfectés et le milieu est changé 4 heures après transfection. Après 24 ou 48 heures le surnageant de culture est récupéré pour être centrifugé et filtré afin de concentrer les vecteurs lentiviraux. Le culot final est repris dans 1 mL de DMEM et sera dosé par la plateforme en mesurant la quantité de p24 par ELISA. Le titre infectieux est également déterminé en infectant des cellules 293T avec des dilutions sérielles de pseudovirus et par la mesure de la moyenne de fluorescence par cytométrie en flux.

- **Transduction**

Les transductions basées sur l'utilisation des vecteurs lentiviraux se fait dans un laboratoire de type P2. Dans un premier temps les cellules sontensemencées dans des plaques 48 puits à raison de 10 000 cellules par puits dans un volume de 400 µL. Après 48h d'incubation à 37°C, les cellules sont lavées au PBS, le milieu de culture est renouvelé et les cellules sont transduites avec différentes multiplicités d'infection (MOI) de vecteurs lentiviraux. Quand les cellules arrivent à confluence, elles sont transvasées dans des plaques 6 puits dans un volume final de 2 mL. L'analyse de la fluorescence est effectuée 10 jours après la transduction par cytométrie en flux. Pour cela les cellules sont centrifugées pendant 5 minutes à 2000 rpm puis le culot est repris dans un mélange PBS-EDTA (0,2 mM) –SVF (0,5%).

d) Infection par le virus sauvage

- **Isolation des PBMC**

Les PBMC sont isolées à partir d'échantillons de sang de patients séronégatifs fournis par l'EFS. On fait dans un premier temps passer le sang dans un filtre à déleucocyter qui est ensuite soumis à un gradient de Ficoll afin de ne récupérer que les PBMC : 15 mL de Ficoll sont mélangés avec 20-25mL du sang récupéré puis centrifugé 15 minutes à 2000 rpm. Le plasma (solution transparente située en haut) est éliminé, ce qui permet d'accéder aux PBMC qui forment un anneau au-dessus de la solution de Ficoll. Cet anneau est récupéré précautionneusement dans un falcon 50mL et complété avec du PBS stérile qsp 50mL puis centrifugé pendant 10 minutes à 1500rpm pour laver les cellules. Le culot ainsi formé est récupéré puis re-suspendu dans 4mL de milieu de culture. Les cellules sont ensuite numérées afin de faire une culture stock de PBMC contenant $100 \cdot 10^6$ cellules qui seront alors activées avec de la PHA pendant 48 heures.

- **Infection**

Toutes les étapes qui impliquent la manipulation du virus sauvage se font dans un laboratoire de type P3. Les PBMC activées sont à nouveau numérées afin d'en récupérer $60 \cdot 10^6$ pour les centrifuger 5 minutes à 1000 rpm. Le culot cellulaire est ensuite infecté avec $6 \cdot 10^6$ particules virales (MOI de 0,1) pendant 3 heures à 37°C. A l'issue de ce temps, les cellules infectées sont centrifugées et le surnagent est éliminé afin de se débarrasser des virus n'ayant pas infecté de cellules. Le culot est ensuite re-suspendu dans 30mL de milieu de culture et 1mL de cellules infectées sont réparties dans chaque puits d'une plaque 24 puits. Les puits sont ensuite complétés avec 1mL de milieu de culture contenant une des molécules à tester, pour un volume final de 2mL par puits. Les cellules sont incubées à 37°C, 5% de CO₂ pendant 24, 48 ou 72h pour analyser la réplication virale.

- **Analyse de la réplication virale**

La charge virale est déterminée à 24, 48 et 72h post-infection. Pour cela, 1mL de surnagent de culture est prélevé pour chacune des conditions en faisant attention de ne pas prélever de cellules. Le nombre de copies d'ARN viral est ensuite déterminé dans ce surnagent par le CHU de Bordeaux avec un cobas 6800.

2.3 Manipulation des acides nucléiques

a) Purification de l'ADN de cellules infectées

L'ADN des cellules infectées est isolé grâce au kit QIAmp DNA Blood Mini Kit (QIAGEN) en suivant le protocole du fournisseur.

b) qPCR

Les qPCR visant à déterminer la quantité d'ADN viral total et intégré sont réalisées en collaboration avec le Dr O.Delelis. Le protocole utilisé ainsi que les amorces sont décrites dans Brussel et Sonigo, 2003¹⁷⁹.

c) Séquençage des sites d'intégration

L'ADN purifié de cellules K562 est analysé sur gel d'agarose 0,7% pour vérifier la qualité et la quantité d'ADN extrait pour chaque condition. Les échantillons sont ensuite fournis à notre collaborateur le Dr Z.Ivics qui s'occupe de la préparation de la librairie en vue du séquençage des sites d'intégration par illumina.

d) Mutagenèse dirigée

La mutagenèse des mutants de l'IN a été réalisée sur le plasmide pET-21b-IN pour l'expression et la purification des IN, et sur le plasmide pCMVΔ8.91 pour la production des vecteurs lentiviraux qui exprimeront une IN mutante. Les mutagenèses ont été réalisées par BioInnovatise Inc.

2.4 Manipulation des protéines

a) Production de l'IN du VIH-1

Afin de réaliser la purification d'IN recombinante WT ainsi que l'ensemble des IN mutantes, une préculture est réalisée par ensemencement d'une colonie de Rosetta transformée avec le pET-21b-IN dans 20 ml de LB + ampicilline (100 µg/ml) + chloramphénicol (34 µg/ml) durant la nuit à 37 °C sous agitation (220 rpm). Puis, la préculture est ensemencée dans 500 ml de LB complété avec de l'ampicilline et du chloramphénicol jusqu'à l'obtention d'une DO à 600nm de 0,6. L'induction de la production intégrase se fait par l'ajout d'IPTG (500 µM). La culture est ensuite incubée 3 heures à 37 °C sous agitation (220 rpm). Les bactéries sont ensuite centrifugées 10 min à 4000 rpm à 20 °C et les culots protéiques sont congelés à 80 °C.

Tampon de lyse : 50 mM HEPES pH 7,5, 5 mM EDTA pH 8, 1 mM DTT, 1 mM PMSF, inhibiteurs de protéases (Roche)

Tampon de solubilisation : 50 mM HEPES pH 7,5, 1 M NaCl, 1mM EDTA, 1mM DTT, 7 mM CHAPS

Tampon LCS (Low salt concentration) : 50 mM HEPES pH 7,5, 0,2 M NaCl, 0,1 mM EDTA, 1mM DTT, 7 mM CHAPS, 10 % glycérol.

Tampon HCS-Butyl (High salt concentration) : 50 mM HEPES pH 7,5, 0,2 M NaCl, 0,1 mM EDTA, 1 mM DTT, 7 mM CHAPS, 1M Sulfate d'ammonium, 10 % glycérol. Tampon HCS-Héparine : 50 mM HEPES pH 7,5, 1 M NaCl, 0,1 mM EDTA, 1 mM DTT, 7 mM CHAPS, 10 % glycérol.

Tampon de dilution : 50 mM HEPES pH 7,5, 0,1 mM EDTA, 1mM DTT, 7 mM CHAPS, 10 % glycérol

Les culots protéiques sont repris dans 20 ml de tampon de lyse complété avec du lysozyme (20 mg/ml) et de la DNase (15 µl/ml) (RQI Rnase free Promega). Après sonication à 60 W pendant 20 secondes, une centrifugation à 10 000 rpm pendant 10 minutes est réalisée. Le surnageant constituant la fraction protéique soluble et le culot est ensuite lavé dans 20 ml de tampon de lyse puis est centrifugé. Le culot est ensuite repris dans 13 ml de tampon de solubilisation pendant 1 heure à 4 °C sous agitation. Puis, le culot solubilisé est centrifugé 1 h à 33 000 rpm. Le surnageant est ensuite supplémenté avec du sulfate d'ammonium à une concentration finale de 1M et agité 5 min à 4 °C. L'ensemble est ensuite centrifugé 20 min à 10 000 rpm à 4 °C et le surnageant obtenu est alors chargé sur colonne HiTrap Butyl Sépharose 4B (1 ml, GE Healthcare) préalablement lavée avec 5 volumes de LSC puis équilibrée avec 5 volumes de HSC. L'élution se fait par un step de NaCl (1 à 0 M). La densité optique des fractions éluées est lue à 280 nm. Les fractions contenant l'IN sont regroupées et diluées au 1/3,3 dans le tampon de dilution. La deuxième colonne est une HiTrap Héparine Sépharose CL-4B (1 ml, GE Healthcare). Cette colonne est premièrement lavée avec 5 volumes de tampon HS puis équilibrée avec 5 volumes de tampon LCS. L'ensemble des fractions diluées est ensuite chargé sur la colonne et les protéines sont éluées par choc ionique de NaCl. Les fractions contenant l'IN sont diluées 4 fois dans du tampon de dilution puis rechargées sur une seconde colonne Héparine Sépharose CL-4B. L'élution de cette dernière colonne se fait par gradient de NaCl (0,2 à 1 M). Les fractions sont analysées par SDS-PAGE 12 % suivie d'une coloration au bleu de Coomassie et la concentration est déterminée au Nanodrop.

b) Purification des intasomes

La purification et l'assemblage des intasomes de Maedi Visna Virus (MVV) et du Prototype Foamy Virus (PFV) ont été réalisés suivant les protocoles décrits dans Ballandras et al., 2017⁹⁰ et Maskell et al., 2015¹²⁷ respectivement.

c) Western Blot

Après la migration des gels de protéines SDS-PAGE, les protéines sont transférées sur une membrane PVDF grâce au dispositif de transfert des gels iBlot de Invitrogen. La révélation

des western blots est réalisée par chimiluminescence grâce au kit Clarity Western ECL Substrate de BIO-RAD et la prise d'image est réalisée avec un appareil ImageQuant LAS 4010.

2.5 Tests in vitro

a) Far Dot Blot

Des dilutions sérielles de l'IN (1µL par dilution) sont déposées sur une membrane de nitrocellulose et laissées à sécher pendant une heure à température ambiante. La membrane est ensuite saturée avec une solution de BSA à 1% diluée dans du PBS pendant 3 heures. Après 2 lavages au PBS la membrane est incubée dans une solution contenant 1µM de peptide biotinylé choisi dilué dans du PBS pendant 1 heure. La membrane est ensuite lavée 2 fois au PBS et incubée avec une solution d'extravidine couplée à la peroxydase (SIGMA) au 1/4000 dans une solution de PBS-0,3% BSA pendant 1 heure. L'interaction est ensuite analysée comme pour un western blot par chimiluminescence.

b) Thermophorèse

Les expériences de thermophorèse ont été réalisées en collaboration avec le Dr M. Ruff. Brièvement, des dilutions en série du domaine CTD de l'IN du VIH-1 sont effectuées puis, le peptide d'intérêt est ajouté au CTD de l'IN. L'échantillon est ensuite incubé 15 min à température ambiante. L'ensemble des mesures est effectué par le Nanotemper Monolith NT.015 avec un temps de « laser-on » de 30 sec et un temps de « laser-off » de 5 sec dans les conditions de 20 % LED et 40 % MST IR-Laser power.

c) Test d'intégration concertée

L'IN purifiée recombinante est diluée à 2µM dans un tampon 1M NaCl ; 20mM Hepes pH7 ; 10mM DTT. L'IN (200nM) est ensuite incubée avec un ADN donneur H55 radiomarké au P³² (10nM) et un plasmide p481 mimant l'ADN receveur (50ng) pendant 30 minutes dans la glace. La réaction est initiée par l'ajout du tampon de réaction (concentrations finales : 17% DMSO ; 6 mM DTT ; 11 mM MgCl₂ ; 22 µM ZnCl₂ ; 8% PEG) pendant 2 heures à 37°C (volume final 10µL) et stoppée par une solution protéinase K (1 mg/ml), SDS (0,7%), EDTA (7 mM) pendant 1 heure à 55°C. Les produits de la réaction sont extraits par un traitement phénol-chloroforme-alcool isoamylique (24/25/1 ; v/v/v) et chargés sur gel d'agarose 1% pour une migration de 2 heures à 200V. Le gel est ensuite traité à l'acide trichloroacétique 5% pendant 20 minutes pour fixer les acides nucléiques, et séché 1 heure à 80°C sous vide. Les produits

d'intégration sont alors révélés par autoradiographie grâce au système PharosFX™ et sont quantifiés avec le logiciel Quantity One.

Les tests d'intégration sur mononucléosomes sont réalisés dans les mêmes conditions mais avec 400nM d'IN, un ADN donneur HU5 et un mononucléosome assemblé sur séquence 601 (250ng) comme receveur. A l'issue de la réaction les produits sont précipités à l'éthanol avant d'être chargés sur un gel de polyacrylamide dénaturant 8%.

d) Pull Down

Les billes magnétiques DynabeadsMyOne Streptavidin T1 (Invitrogen) sont dans un premier temps lavées (12,5µL par condition) avec un tampon d'interaction (50mM Hepes pH7,5 ; BSA 0,1mg/mL ; DTT 1mM ; Tween20 0,1% ; Glycérol 10% ; 200mM NaCl). Les billes lavées sont ensuite incubées avec l'IN recombinante (10pmoles) ainsi que l'ADN 601 (125ng) ou les mononucléosomes biotinylés (250ng, équivalant à 125ng d'ADN 601) pendant 1h à 4°C sous rotation dans un volume final de 300µL (pour les expériences nécessitant du peptide, les versions non biotinylées de ces derniers ont été utilisées pour les expériences de pull down). Les billes sont magnétisées et les protéines qui n'ont pas interagi sont éliminées en jetant le surnageant. Les billes sont lavées deux fois avec 300µL de tampon d'interaction et les produits précipités avec les billes sont re-suspendus dans du tampon de charge de protéines et chargés sur gel SDS-PAGE. L'interaction sera finalement révélée par western blot anti-IN ou anti-tag selon l'IN utilisée.

e) Test de 3' processing et de transfert de brins

Les tests de 3' processing et de transfert de brins sont réalisés dans les mêmes conditions que les tests d'intégration sur mononucléosomes sauf que l'ODN utilisé (hybride entre l'ODN 70 et 72 pour le 3' processing et l'hybride 70-71 pour le test de transfert de brins) est à 20nM final.

f) Test d'intégration avec les intasomes

Les tests d'intégration utilisant les intasomes purifiés ont été réalisés avec 200ng du plasmide p481 ; 7,5mM MgSO₄ et 10nm d'intasome purifié dans un volume final de 40µL. Après incubation à 37°C pendant 5 minutes les produits sont précipités à l'éthanol et repris dans du tampon de charge ADN pour être chargés sur gel d'agarose 0,8%. Les produits d'intégration sont ensuite révélés au SyberSafe.

g) Pontages chimiques au BS3

L'IN recombinante est diluée à 5µM dans un tampon (1M NaCl ; 20mM Hepes pH7 ; 400µM ZnCl₂ and 10mM DTT) afin que l'IN ne soit pas agrégée avant le pontage. L'IN (1µL) est ensuite incubée avec le peptide choisi pendant 30 minutes sur la glace. Les échantillons sont ensuite pontés avec une solution de BS3 (concentration finale 20mg/mL) dans un volume final de 10µL pendant 30 minutes à 25°C. La réaction est stoppée par l'ajout du tampon de charge pour protéines et les échantillons sont chargés sur un gel SDS-PAGE et révélés par western blot anti-IN (pour VIH) ou anti-HIS (pour PFV). L'oligomérisation est ensuite quantifiée en mesurant la quantité de complexes protéiques de plus haut poids moléculaires que le monomère avec le logiciel imageJ.

2.6 Analyse *in silico*

a) Modélisation de l'interaction entre le CTD de l'IN et le peptide H4K20me1

La modélisation a été réalisée par nos collaborateurs P.Gouet et X.Robert. Dans l'ensemble des expériences de docking moléculaire, le domaine CTD de l'IN du VIH-1 correspondant au domaine 210-270 issu de la structure cristallographique (PDB 1EX4) a été utilisé comme récepteur. Le pentapeptide H4K20me1 dérivé de la structure du complexe MSL3/H4K20me1 (PDB 3OA6) a été utilisé comme ligand. Puis, le docking du récepteur et du ligand a été analysé avec AutoDockTools. Dans les premières expériences de docking utilisant AutoDock et AutoDock Vina, la surface du récepteur a été traitée comme rigide.

b) Criblage virtuel

Le criblage virtuel a été réalisé en collaboration avec X.Robert et P.Gouet, avec comme cible le CTD de l'intégrase du VIH (code PDB : 1EX4). Le sous-ensemble 'Leads-now' de la chimiothèque virtuelle ZINC12 (<http://zinc.docking.org>) a été utilisé comme source de petites molécules candidates ; ces dernières ont été préparées avec le logiciel AutoDockTools. Les 3,687,621 composés de la chimiothèque ont alors été testés grâce au logiciel Autodock/Vina sur la grappe de calcul du laboratoire "Rétrovirus et Biochimie Structurale" (UMR5086, IBCP, Lyon). Dans cette première étape, la protéine cible a été traitée comme rigide et les composés comme flexibles. En conservant les meilleurs scores de docking, ceci nous a conduit à isoler environ 300,000 composés qui ont été soumis à une deuxième phase de docking, cette fois-ci en considérant la cible protéique comme partiellement flexible au niveau de la zone d'interaction considérée. Pour ce faire, le logiciel Autodock/Vina a été à nouveau employé, ce qui a permis d'isoler au final 10 molécules candidates pouvant venir se fixer au niveau de la cavité où le peptide est sensé interagir.

RESULTATS

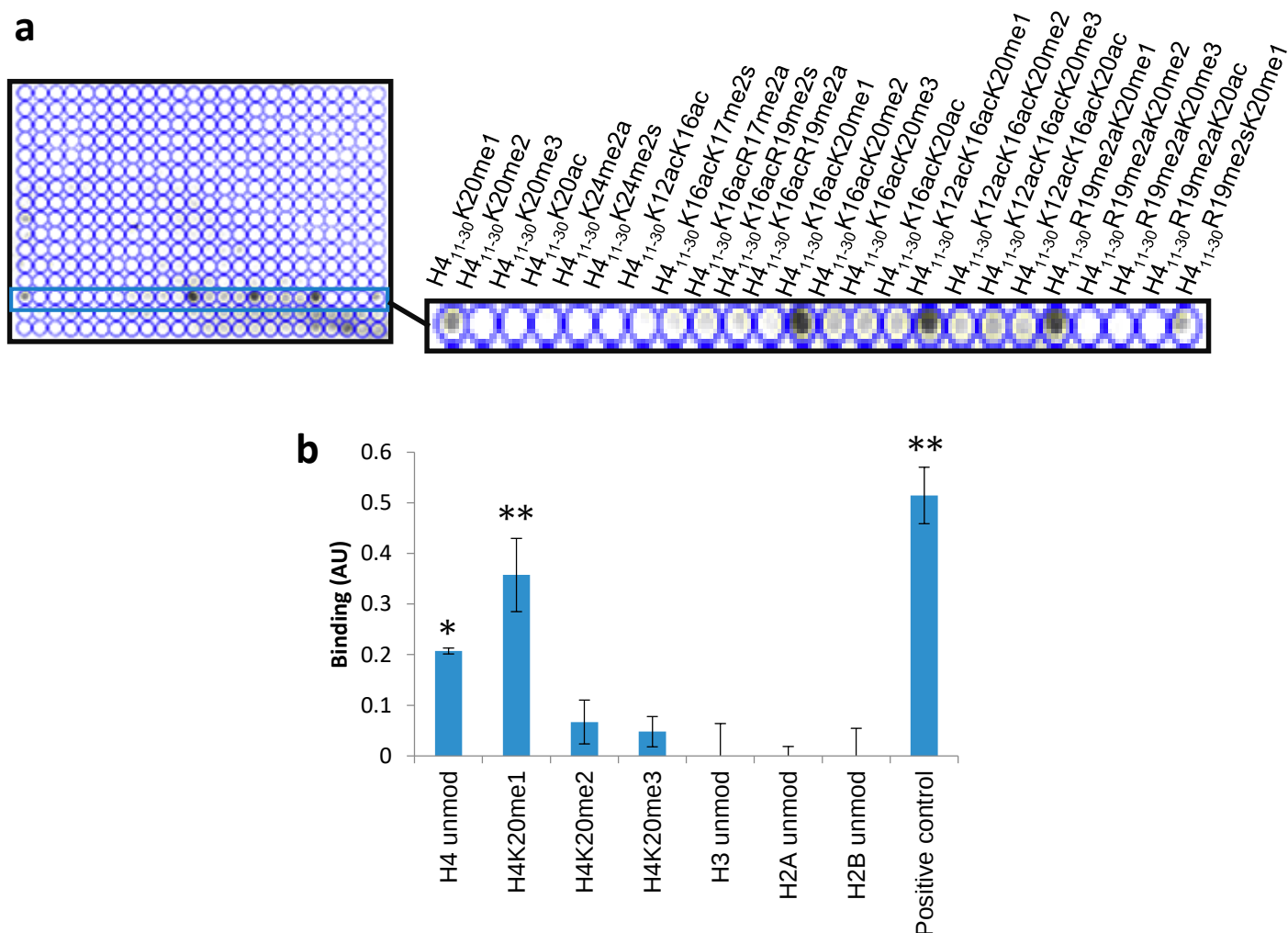


Figure 32 : Interaction entre l'IN du VIH-1 et des peptides dérivés de modifications d'histones. (a) Criblage d'interactions potentielles entre l'IN du VIH-1 et divers peptides mimant des modifications d'histones par histone peptide array. **(b)** Quantification de l'interaction entre l'IN du VIH-1 et des peptides d'histones dérivés de la queue H4 ainsi que des autres queues. Les écarts types ont été calculés avec un test t de Student et représentés par * $p < 0,05$ et ** $p < 0,005$ pour représenter la significativité par rapport aux contrôles négatifs.

Partie 1 : L'interaction entre l'IN du VIH-1 et la queue d'histone H4 est fonctionnelle

Le travail présenté ici constitue une partie des résultats publiés dans : « **Modulation of the functional association between the HIV-1 intasome and the nucleosome by histone amino-terminal tails** »

Mohamed S. Benleulmi & Julien Matysiak & Xavier Robert & Csaba Miskey, Eric Mauro, Delphine Lapaillerie, Paul Lesbats, Stéphane Chaignepain, Daniel R. Henriquez, Christina Calmels, Oyindamola Oladosu, Eloïse Thierry, Oscar Leon, Marc Lavigne, Marie-Line Andreola, Olivier Delelis, Zoltán Ivics, Marc Ruff, Patrice Gouet and Vincent Parissi, *Retrovirology*, 2015.

1 Contexte scientifique et résultats préliminaires

La sélectivité d'intégration du VIH-1 est un processus déterminant pour le bon déroulement du cycle viral car cela définit le devenir des gènes intégrés et notamment leur expression ou leur latence. Pour le cas du VIH-1, les facteurs impliqués dans la sélectivité de ce rétrovirus sont maintenant assez bien décrits (voir la partie 4 de l'introduction pour plus de détails). En résumé, une fois que le CPI a atteint les pores nucléaires, CPSF6 en interagissant avec la capsid du virus, va permettre le transport actif du CPI dans le noyau et va le guider dans des zones peu condensées de la chromatine. Ensuite LEDGF/p75 interagit avec à la fois l'IN et des nucléosomes portant la modification d'histone H3K36me3 présents dans les unités de transcription actives des gènes. Cependant, la déplétion de CPSF6 et/ou de LEDGF n'entraîne pas une perte totale des sites d'intégration dans ces régions, ce qui suggère que d'autres facteurs participent également à ce ciblage. Le laboratoire a montré que *in vitro* et *in vivo*, la condensation de la chromatine inhibait l'intégration du VIH suggérant que des contacts entre l'intasome et le nucléosome pouvaient être masqués dans le cas d'une chromatine compacte. C'est pour ces raisons que le laboratoire s'est demandé si l'IN avait la capacité d'interagir directement avec les protéines constituant le nucléosome que sont les histones. Les queues d'histones apparaissent comme de bons candidats, car elles sont exposées hors de l'ADN nucléosomal et sont facilement accessibles pour interagir avec d'autres protéines.

Pour mettre en évidence de potentielles interactions entre l'IN du VIH-1 et les queues d'histones, l'équipe a criblé 386 combinaisons de modifications d'histones dérivées des queues H2A, H2B, H3 et H4 pour leur interaction avec l'IN par « histone array » (**Figure 32a**).

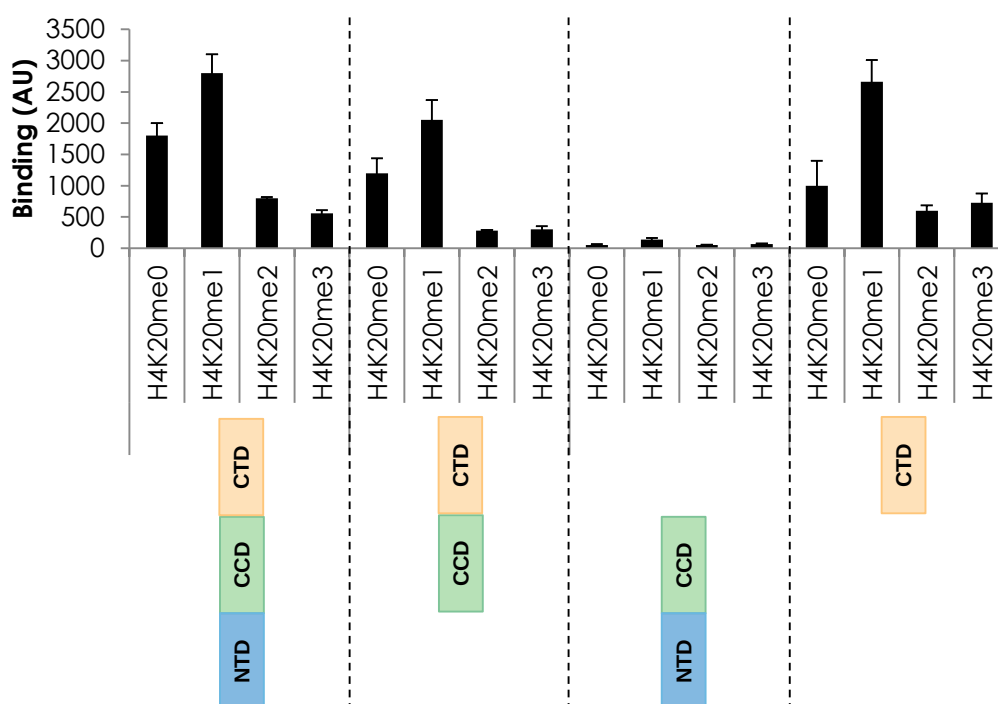


Figure 33 : Identification de la région de l'IN du VIH-1 interagissant avec les peptides dérivés de la queue d'histone H4. L'interaction entre l'IN entière ou différentes formes tronquées avec les peptides dérivés de la queue d'histone H4 a été quantifiée par far dot blot. L'IN entière ou tronquée est fixée sur une membrane de nitrocellulose puis incubée avec un des peptides (H4K20me0, H4K20me1, H4K20me2 ou H4K20me3) qui sont couplés à la biotine. L'interaction est révélée par ajout de peroxydase couplée à la streptavidine et par chimiluminescence puis le signal résultant est quantifié.

Avec cette technique il a été montré que l'IN avait la capacité d'interagir avec des peptides issus de la queue H4 (**Figure 32b**). L'analyse de ces données a également permis de montrer que l'état de méthylation de la lysine 20 module l'interaction avec l'IN. En effet, l'analyse de l'interaction des différentes formes méthylées de H4K20 montre des différences significatives (**Figure 32b**) avec le H4K20me1 restant le meilleur interactant suivi de près par la forme non méthylée. Les peptides H4K20me2 et me3 interagissent toujours avec l'IN, mais de manière bien moins importante. Certaines modifications secondaires en plus du H4K20me1 peuvent augmenter l'interaction, comme le H4K16acK20me1 ou le H4R19me2K20me1. De manière intéressante, la modification d'histone H4K20me1 est trouvée enrichie au niveau des sites d'intégration du VIH-1, laissant supposer qu'elle puisse être impliquée dans la sélectivité. Cette modification H4K20me1 étant le motif minimal requis pour une bonne interaction avec l'IN, c'est sur cette modification que s'est focalisé ce projet de thèse.

2 Identification de la zone d'interaction de l'IN avec la queue d'histone H4

Afin de mieux caractériser l'interaction entre l'IN du VIH-1 et la queue d'histone H4K20me1, nous avons dans un premier temps identifié la zone de l'IN impliquée dans cette interaction. Pour cela, nous avons purifié différentes constructions d'IN comportant des domaines isolés (NTD, CCD ou CTD) ou associés deux à deux (NTD-CCD ou CCD-CTD). Ces constructions ont été testées par far dot blot pour leur capacité à interagir avec un peptide dérivé de la queue d'histone H4 (résidus 11 à 23) non modifié ou portant divers degrés de méthylation sur la lysine 20. Comme montré sur la **Figure 33**, la quantification des far dot blots nous a permis d'observer que la délétion du domaine NTD n'affecte pas l'interaction, suggérant que ce domaine n'est pas important pour l'association avec les histones. En revanche, la délétion du domaine CTD entraîne une perte d'interaction avec les histones, suggérant que ce domaine est requis pour l'interaction. Ceci a été confirmé avec le CTD seul, dont le profil d'interaction avec les histones est très proche de celui de l'enzyme entière, suggérant que ce domaine CTD puisse posséder une fonction de liaison avec les histones.

Dans l'optique de quantifier la force d'interaction entre ces différents partenaires, nous avons réalisé en collaboration avec le Dr M. Ruff, des expériences de thermophorèse. Cette méthode permet l'obtention de constantes de dissociation (Kd) en solution afin de mesurer l'affinité des partenaires entre eux. Brièvement, les mouvements des composés en solution sont analysés et comparés quand ils sont seuls ou quand les deux partenaires sont présents. L'interaction entraîne la formation de complexes de tailles plus importantes, modifiant ainsi leur mouvement en solution. Les différences de comportement de ces complexes à différentes températures permettront l'évaluation de leur Kd. Ainsi, cette étude nous a permis de calculer le Kd entre le CTD de l'IN et les différents peptides issus de la queue d'histone H4K20 (**Figure 34**). Bien que dans ces conditions nous n'avons pas détecté d'interaction avec le H4K20me0, l'association avec le H4K20me1 a pu être confirmée avec un

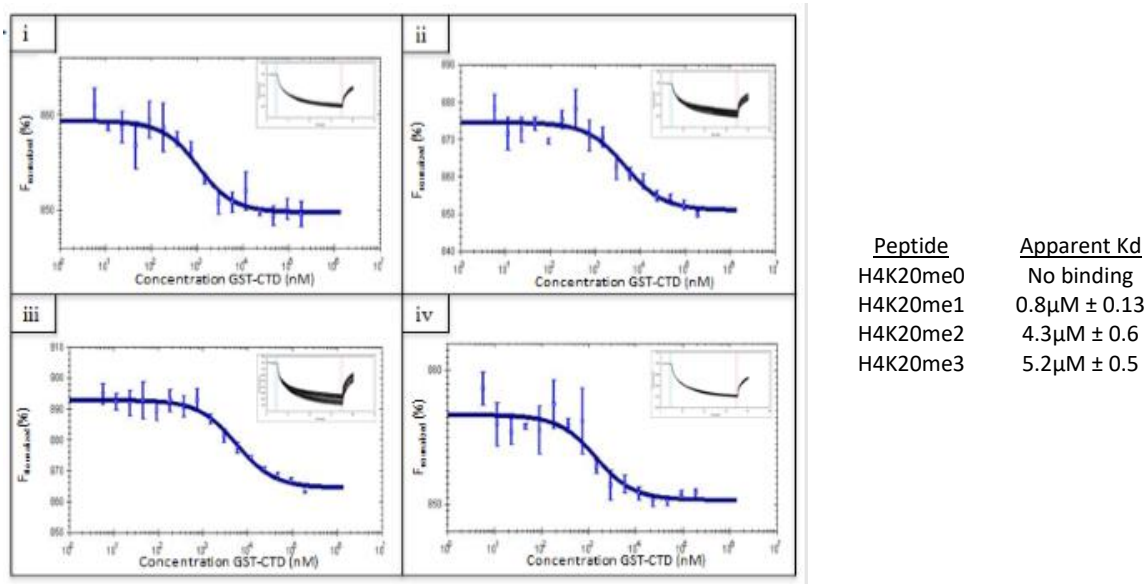


Figure 34 : Détermination de la constante de dissociation (Kd) entre le CTD de l'IN du VIH-1 et les peptides dérivés de la queue d'histone H4. L'analyse quantitative de l'interaction entre le CTD de l'IN du VIH et les peptides dérivés de la queue d'histone H4 (H4K20me0, H4K20me1, H4K20me2 ou H4K20me3) a été effectuée par thermophorèse. Les compétitions ont été effectuées par l'ajout du peptide H4K20me0 et H4K20me1. Les graphiques représentent l'intensité de fluorescence normalisée en fonction de la concentration de CTD. Les valeurs de Kd déduites de ces expériences sont indiquées sur la droite des graphiques.

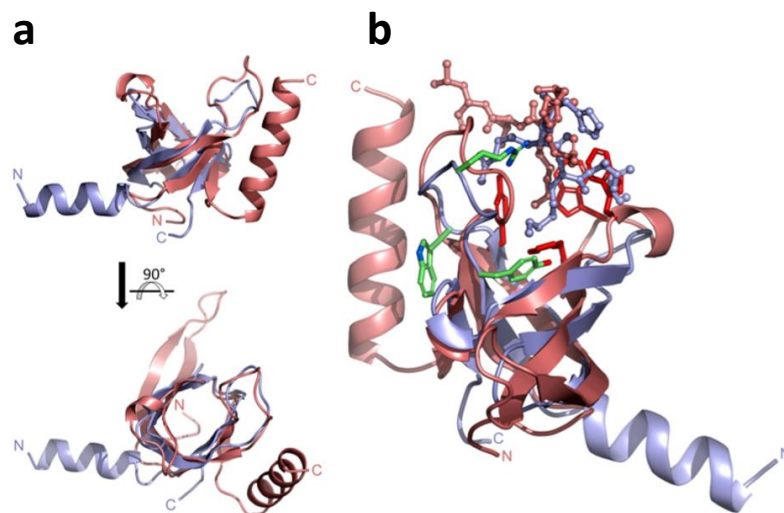


Figure 35 : Comparaison de structure entre le CTD de l'IN du VIH-1 et la protéine hMSL3. La structure RMN du CTD de l'IN du VIH-1 (en bleu) et la structure cristallographique de hMSL3 (en rouge) sont représentées en cartoon. **(a)** Superposition de la structure tridimensionnelle des deux protéines vue sous deux orientations. **(b)** Superposition de la modélisation d'interaction du CTD-peptide H4K20me1 avec la structure cristallographique du complexe hMSL3-peptide H4K20me1. Dans les deux cas le peptide est représenté en boules et bâtons.

Kd d'environ 0,8 μ M. L'affinité obtenue pour le H4K20me2 et H4K20me3 est plus faible que la forme monométhylée, avec respectivement des Kd d'environ 4,3 et 5,2 μ M. Ces données suggèrent que l'IN du VIH-1 est capable d'interagir directement avec un peptide issu de la queue d'histone H4 *in vitro*, et que cette interaction se fait via son domaine CTD. Pour confirmer nos hypothèses, nous nous sommes intéressés de plus près à ce domaine C-terminal afin d'identifier plus précisément un motif ou des résidus impliqués dans cette interaction avec la queue d'histone H4.

3 Analyse structurale et modélisation moléculaire *in silico* de l'interaction entre le CTD de l'IN et la queue d'histone H4

De manière intéressante, le CTD de l'IN du VIH présente un repliement proche de celui d'un motif SH3, connu pour être impliqué dans les interactions protéine-protéine et plus particulièrement pour reconnaître des lysines ou des arginines méthylées^{86,180}. Le motif SH3 est retrouvé en particulier chez des protéines ayant des domaines appartenant à la « royal family » tels que les chromodomaines, domaines tudor, MBP ou PWWP¹³⁷. Chacun de ces domaines est capable de reconnaître des modifications épigénétiques d'histones. Nous nous sommes donc demandé si un tel motif dans le CTD de l'IN pouvait être fonctionnel pour la reconnaissance de ces modifications. Nous avons pour cela comparé la structure du CTD de l'IN avec hMSL3, un facteur de transcription possédant un chromodomaine capable de reconnaître la modification H4K20me1 spécifiquement^{181,182}. De manière frappante, la superposition de ces deux structures montre une très forte homologie structurale entre ces deux protéines (**Figure 35**). Bien que le CTD de l'IN ne possède pas toutes les caractéristiques structurales d'un chromodomaine, ces observations suggèrent tout de même qu'il puisse fonctionner comme un lecteur de la chromatine en reconnaissant des modifications d'histones telle que la marque H4K20me1.

Pour mieux caractériser la zone d'interaction IN-queue d'histone H4, nous avons voulu identifier les résidus du CTD qui sont importants pour la liaison avec le peptide H4K20me1. Pour cela, une modélisation moléculaire *in silico* de ces deux partenaires a été effectuée en collaboration avec le Dr P.Gouet et X.Robert, afin d'identifier les zones possibles où un pentapeptide mimant les résidus 18 à 22 de la queue d'histone H4 peut se fixer sur le CTD. Cette analyse a été réalisée en utilisant deux programmes différents : AutoDock et AutoDock Vina. Une seule zone d'interaction commune a été trouvée grâce à ces programmes (classée 2^{ème} avec AutoDock et 1^{ère} avec AutoDock Vina). Cette zone en forme de V est délimitée par les boucles 228-235 et 253-257 (**Figure 36**). Dans cette modélisation, on peut voir le peptide qui vient se fixer dans cette région en forme de V, avec la lysine 20 pointant au fond de cette cavité. Si le peptide H4 se fixe bel et bien à cet endroit, nous pouvons supposer que l'espace restreint autour de la lysine 20 ne laisse pas la place pour la di ou tri méthylation de cette position, ce qui pourrait expliquer la plus faible affinité du H4K20me2 et H4K20me3 pour le CTD de l'IN. A partir de cette modélisation, les résidus de l'IN qui semblent impliqués dans

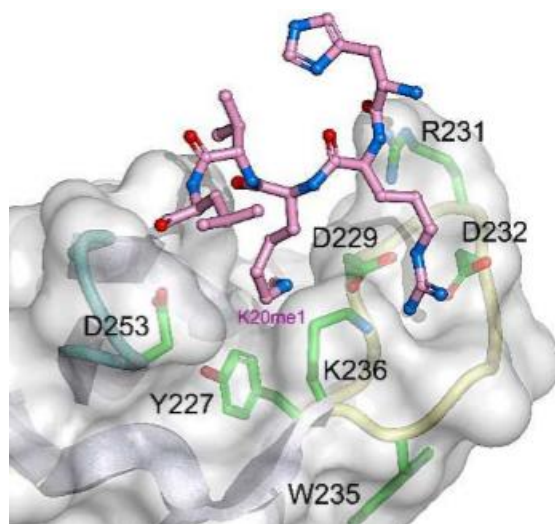


Figure 36 : Modélisation de l'interaction entre le CTD de l'IN du VIH-1 et un pentapeptide dérivé de l'histone H4K20me1. Représentation en surface du CTD de l'IN du VIH-1 en gris et d'un pentapeptide dérivé de H4K20me1 en violet représenté en boules et bâtons. La boucle 228-235 est colorée en jaune et la boucle 253-257 en cyan. Les résidus d'intérêt Y227, D229, R231, D232, W235, K236 et D253 sont représentés en bâton de couleur verte. La lysine 20 de la queue d'histone H4 est également indiquée comme pointant vers le fond de la cavité en forme de V que forme cette région de l'IN.

l'interaction avec le peptide ont été identifiés ainsi que la nature de ces interactions. C'est ainsi que nous avons pu observer que le peptide pouvait potentiellement engager 7 liaisons hydrogènes avec les résidus D229, R231, S255, D256 et K258 ainsi que 15 liaisons hydrophobes avec les résidus Y227, D229, N232, K236, D256, K258 et V260. Dans ce modèle, la méthylation de la lysine 20 à elle seule engage 7 des 15 liaisons hydrophobes et 1 des 7 liaisons hydrogènes avec le CTD, montrant à quel point cette modification semble cruciale dans cette interaction.

4 Recherche des résidus du CTD impliqués dans l'interaction avec la queue d'histone H4

Pour confirmer notre modèle et développer des outils pour mieux caractériser l'interaction entre l'IN et la queue d'histone H4, nous nous sommes orientés vers la mutagenèse dirigée des résidus du CTD de l'IN qui semblent importants pour l'interaction avec le peptide. Ceci nous permettrait d'une part de conforter notre modèle et la zone d'interaction que nous avons identifiée par modélisation, mais également d'obtenir des IN mutantes étant altérées pour leur interaction avec la queue d'histone, afin de les étudier dans des tests fonctionnels. Le CTD de l'IN étant impliqué dans de multiples fonctions au cours du cycle viral (interaction avec la RT, interaction avec l'ADN cible du nucléosome, oligomérisation, etc...), certaines positions ont été délicates à muter sans altérer l'activité de l'enzyme. Le choix des positions à muter s'est porté sur les liaisons hydrogènes et hydrophobes identifiées précédemment, ainsi que sur des homologies structurales avec hMSL3 et les chromodomaines. C'est ainsi que nous avons choisi de muter les positions Y227, D229, R231, W235, K236 et D253. Le Y227 est au fond de la cavité en forme de V et se trouve à une position importante pour la formation de la cage aromatique des chromodomaines, en plus d'être impliqué dans de nombreuses liaisons avec le K20me1 du peptide H4. D229 et K236 engagent des liaisons avec le K20 et le R19 du peptide H4. Le R231 se trouve au sommet de la cavité et semble important pour la stabilisation de l'extrémité N-terminale du pentapeptide en interagissant avec le résidu H18. Le W235 n'est à priori pas impliqué dans l'interaction directe avec le peptide, cependant il est localisé dans la boucle 228-236 et semble important pour l'architecture de la cavité. De même pour le D253, qui ne semble pas interagir directement avec le peptide, mais il fait partie de la boucle 253-257 de l'autre côté de la cavité participant également à sa structuration. Ces positions ont été mutées en alanine, glycine ou histidine dans l'optique d'obtenir des mutants dont l'interaction avec la queue d'histone H4 est modifiée sans toutefois altérer l'activité catalytique de l'enzyme pour éviter tout biais pour la suite de l'étude. Nous avons pour cela choisi des mutations qui vont empêcher les acides aminés identifiés précédemment d'engager des interactions hydrophobes ou hydrogènes avec le peptide, mais qui ne changent pas fondamentalement la nature de cet acide aminé pour limiter au plus les changements structuraux que la mutagenèse pourrait entraîner sur le CTD. Les IN mutantes ont ensuite été testées pour leur capacité à interagir avec le peptide H4K20me1. Le polymorphe naturel D232G a également été inclus dans l'étude.

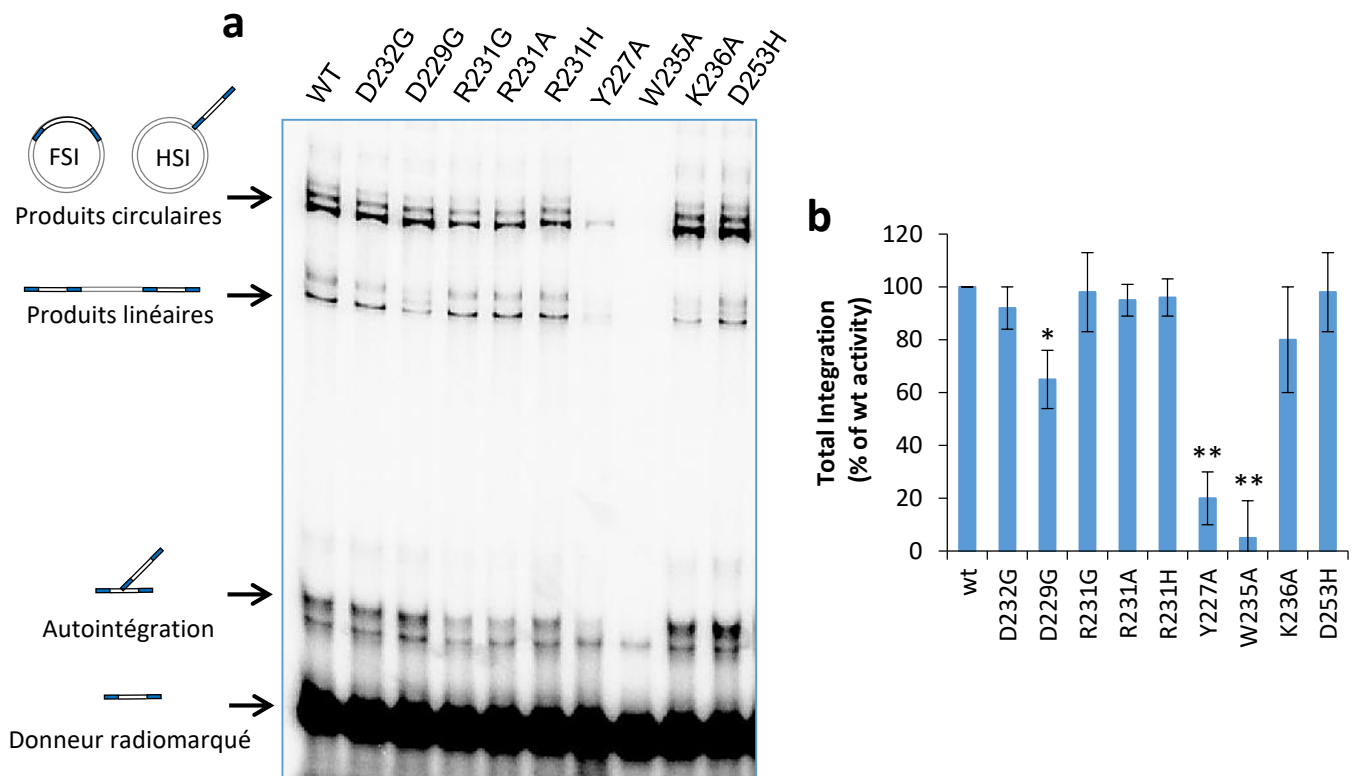


Figure 37 : Activité catalytique des différents mutants de l'IN sur ADN nu. (a) L'activité intrinsèque des différentes IN recombinantes purifiées au laboratoire a été analysée par un test d'intégration concertée réalisée sur ADN nu. Les différents produits d'intégration sont schématisés sur la gauche du gel. **(b)** Quantification de l'intégration (formes circulaires et linéaires) des différents mutants de l'IN normalisée par rapport à l'activité de l'enzyme sauvage. Les écarts types ont été calculés avec un test t de Student et représentés par * $p < 0,05$ et ** $p < 0,005$ sur au moins trois expériences indépendantes.

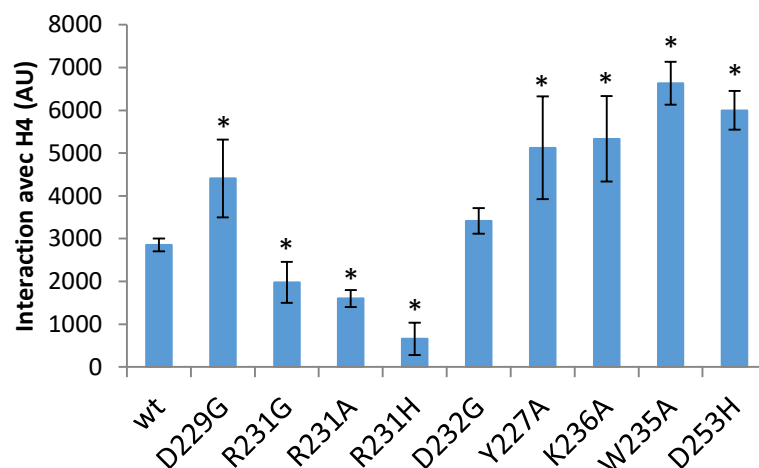


Figure 38 : Quantification de l'interaction entre les mutants de l'IN et le peptide dérivé de la queue d'histone H4K20me1 par far dot blot. Les IN mutantes ont été fixées sur une membrane de nitrocellulose et incubées avec un peptide dérivé de la queue d'histone H4K20me1 biotinylé. L'interaction est révélée par ajout de peroxydase couplée à la streptavidine et par chimiluminescence puis le signal résultant est quantifié. Les écarts types ont été calculés avec un test t de Student et représentés par * $p < 0,05$ et ** $p < 0,005$ sur trois à dix expériences indépendantes.

Après purification de ces intégrases nous avons vérifié leur activité catalytique *in vitro* à l'aide d'un test d'intégration concertée sur plasmide. En effet nous devons nous assurer que les mutants étaient toujours actifs afin éviter tout biais pour la suite de l'étude. Comme on peut le voir en **Figure 37**, certaines mutations entraînent une perte drastique d'activité comme observé pour les IN Y227A et W235A qui ne conservent pas plus de 25% d'activité comparé à l'enzyme sauvage. Le D229G et le K236A sont également impactés mais de manière moins drastique, ils conservent environ 70-80% d'activité. En revanche les mutants R231A/G/H et D253H sont aussi actifs que l'enzyme sauvage, et ce sont donc sur ces positions que nous nous sommes focalisés pour la suite de l'étude.

Les mutants ont ensuite été testés pour leur capacité à interagir avec le peptide H4K20me1 grâce à des expériences de far dot blot. Pour cette analyse, nous avons gardé l'ensemble des mutants que nous avons générés pour connaître leur profil d'interaction avec la queue d'histone. Les résultats des far dot blots réalisés avec les mutants et le peptide H4K20me1 ont été quantifiés (**Figure 38**) et nous avons observé différents comportements de ces mutants : les mutants R231G/A/H présentent une baisse d'affinité pour la queue d'histone H4 (perte d'environ 80% pour la substitution R231H comparé à l'enzyme sauvage), alors que les mutants D229G, Y227A, K236A, W235A et D253H présentent une meilleure affinité pour la queue d'histone. Dans ces conditions, seule l'affinité du variant naturel D232G a été trouvée identique à l'enzyme sauvage. L'ensemble de ces résultats semble corroborer avec la modélisation de l'interaction et suggère fortement que la région identifiée précédemment est bel est bien impliquée dans l'interaction avec la queue d'histone H4 *in vitro*. Les positions que nous avons sélectionnées pour la mutagenèse dirigée sont donc impliquées directement ou indirectement dans l'interaction avec l'histone H4 *in vitro*, ainsi ces mutants vont nous permettre d'étudier la fonctionnalité de l'interaction IN-queue d'histone H4 durant l'étape d'intégration. Pour la suite de l'étude, nous ne retiendrons que les mutants qui sont catalytiquement actifs afin d'éviter tout biais dans l'analyse des résultats (les mutants R231A/G/H qui ont une baisse d'affinité et le mutant D253H qui a une meilleure affinité pour la queue d'histone).

5 Impact des mutations du CTD affectant l'interaction avec l'histone H4 sur les phases précoces de la réplication

L'objectif à présent est d'identifier si l'interaction entre l'intégrase et la queue d'histone H4 est importante pour le cycle de réplication du virus. Pour cela, nous avons introduit les mutations décrites précédemment dans la séquence codante pour l'IN de vecteurs lentiviraux, et après production de ces vecteurs par la plateforme de vectorologie de Bordeaux, nous avons comparé leur infectivité avec celle du vecteur exprimant une intégrase sauvage. Ces vecteurs sont pseudotypés avec la membrane du VSV-G et, mise à part l'entrée, ils récapitulent les étapes précoces de la réplication virale (jusqu'à l'intégration du provirus). Ils possèdent un gène rapporteur codant pour l'eGFP qui permet le suivi de l'infectivité grâce

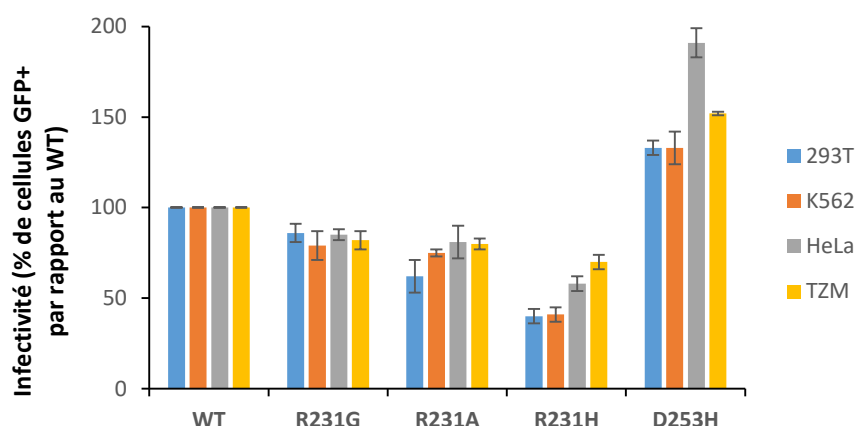


Figure 39 : Infectivité des vecteurs exprimant une IN mutante dans différents types cellulaires. Les différents types cellulaires ont été transduits avec les vecteurs exprimant des IN mutées (R231G/A/H ou D253H). L'infectivité de ces différents mutants a été mesurée par cytométrie en flux en quantifiant le nombre de cellules GFP positives 10 jours après la transduction.

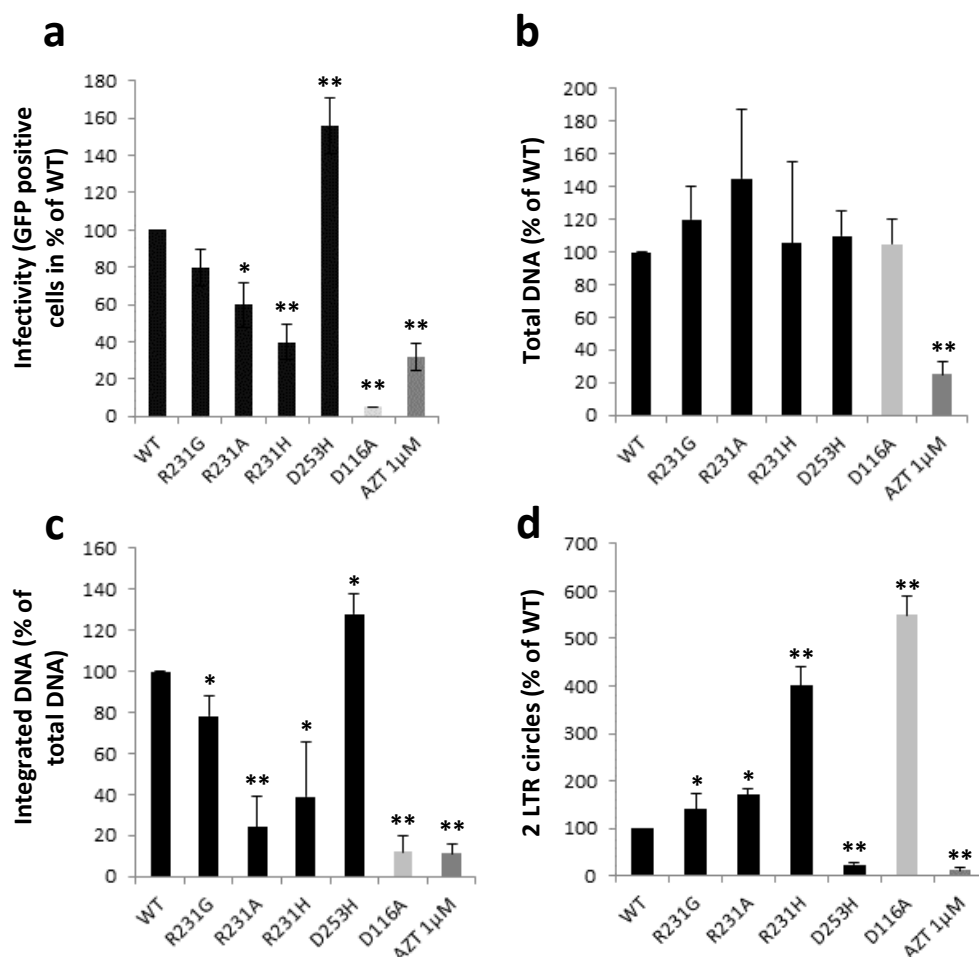


Figure 40 : Effet des mutations de l'IN modulant l'interaction avec le peptide d'histone H4K20me1 sur les phases précoces de la réplication virale. Des cellules 293T ont été transduites par les différents vecteurs lentiviraux. **(a)** L'ensemble des phases précoces de la réplication virale est quantifié par cytométrie en flux. **(b)** La mesure de la quantité d'ADN total **(b)**, d'ADN intégré **(c)** ainsi que les cercles à 2 LTR **(d)** est réalisée par qPCR suite à l'extraction de l'ADN des cellules. Les écarts types ont été calculés avec un test t de Student et représentés par * $p < 0,05$ et ** $p < 0,005$ sur au moins trois expériences indépendantes.

à la quantification de l'expression de l'eGRP par cytométrie en flux. Plusieurs types cellulaires ont été utilisés pour ces expériences (293T, K562, HeLa et TZM) pour s'assurer que les données ne diffèrent pas en fonction du type cellulaire (**Figure 39**).

Les résultats de l'infectivité des différents vecteurs sont présentés en **Figure 40a**, on peut voir que les trois mutants de la position R231 ont tous une infectivité plus faible que les vecteurs ayant une intégrase sauvage, avec une baisse de plus de 60% dans le cas du R231H, à l'inverse du mutant D253H qui lui a une infectivité plus importante (environ 150% du sauvage), (le mutant D116A est connu pour être déficient pour l'étape d'intégration et nous sert ici de contrôle). A ce stade, l'étape du cycle affectée par l'introduction des mutations n'est pas identifiée. Les différentes populations d'ADN virales ont alors été quantifiées par qPCR en collaboration avec le Dr O. Delelis, ENS Cachan. Comme indiqué sur la **Figure 40b**, nous n'avons pas détecté de différence significative dans la quantification de l'ADN total pour aucun des mutants, alors que le contrôle AZT (un inhibiteur de la transcription inverse) diminue bien la quantité d'ADN total. Un défaut dans la transcription inverse a donc été exclu pour ces mutants. Cependant, la quantification de l'ADN intégré (**Figure 40c**) montre qu'il y a des différences entre les mutants : les mutants R231G/A/H présentent une baisse d'environ 25, 80 et 60% respectivement avec une accumulation classique des cercles à 2 LTR (formes d'ADN non intégrées) indiquant un import nucléaire normal (**Figure 40d**). Le D253H a un phénotype opposé, ce mutant possède plus d'ADN intégré que le vecteur sauvage (environ 30% de plus) et moins de cercles à 2 LTR indiquant que l'étape d'intégration est plus efficace pour ce mutant. Bien que nous n'avons pas la preuve directe que ces différences d'intégration soient liées à la capacité d'interaction des intégrases avec la queue H4, la corrélation entre nos données *in vitro* et *in vivo* est tout de même frappante : les trois mutants de la position R231 sont affectés négativement pour l'interaction *in vitro* avec la queue d'histone ainsi que pour leur capacité à s'intégrer *in vivo* dans les cellules, avec une tendance à peu près respectée où le R231H est toujours le plus affecté suivi par le R231A et le R231G. Même corrélation mais avec les phénotypes opposés pour le D253H qui est affecté positivement pour l'interaction *in vitro* et qui s'intègre encore mieux que les vecteurs exprimant une enzyme sauvage. Il semble par conséquent qu'il y ait une très forte corrélation entre la capacité d'interaction des enzymes avec la queue d'histone H4K20me1 et leur capacité à s'intégrer *in vivo*. Pour aller plus loin dans l'analyse de l'effet de ces mutations, nous nous sommes ensuite intéressés à la distribution de leurs sites d'intégration.

6 Impact des mutations du CTD affectant l'interaction avec la queue d'histone H4 sur la sélectivité d'intégration

Pour cette analyse de sélectivité, les cellules K562 ont été utilisées car la cartographie des modifications d'histones et des caractéristiques génomiques sont très bien annotées dans ce type cellulaire. Après transduction des différents mutants dans cette lignée, l'ADN a été extrait trois jours après et séquencé en collaboration avec le Dr Z. Ivics (PEI, Langen). Nous

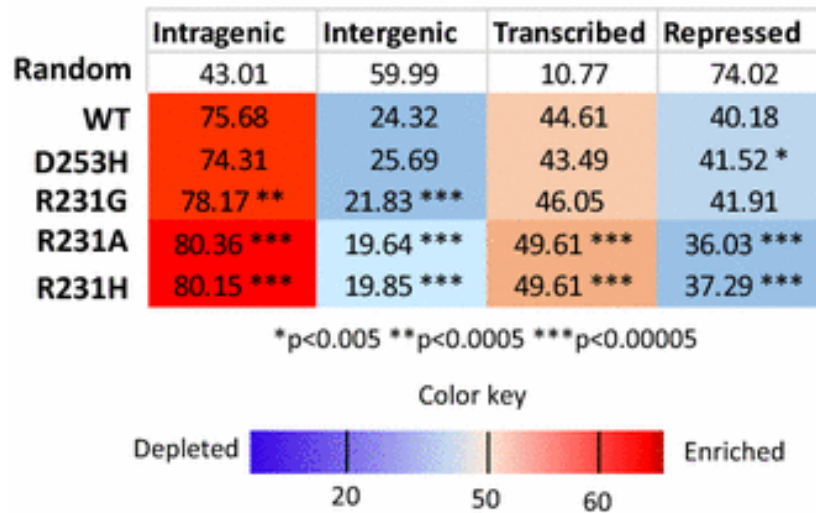


Figure 41 : Analyse de la distribution des sites d'intégration des vecteurs codant pour une IN sauvage ou mutée pour l'interaction avec le peptide histone H4K20me1. Des cellules K562 ont été transduites par les vecteurs lentiviraux pseudotypés VSVG codant pour une IN sauvage ou mutée (R231A/G/H et D253H) puis l'ADN a été récupéré pour séquencer les sites d'intégrations afin de les positionner par rapport aux différentes zones de la chromatine (zone transcrite ou répressive ; zone génique ou intergénique). Les nombres indiquent le pourcentage de sites d'intégration par condition. Le code couleur traduit une déplétion ou un enrichissement des sites d'intégration par rapport à la fréquence aléatoire attendue.

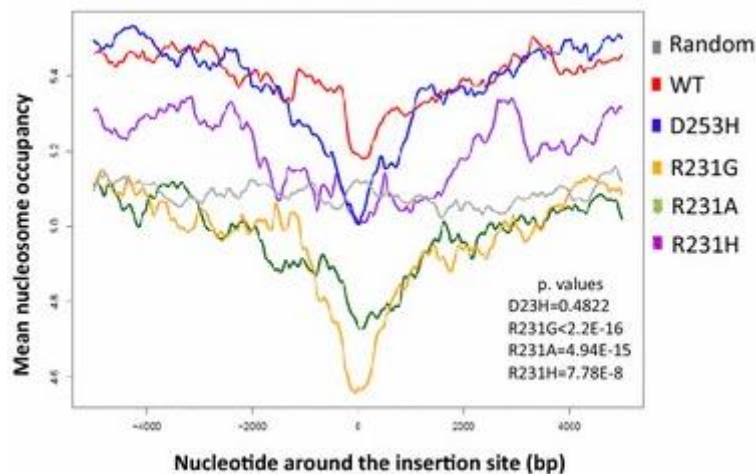


Figure 42 : Analyse de la densité en nucléosomes autour des sites d'intégration des différents vecteurs. Les cartes de la densité en nucléosomes ont été générées à partir des résultats de séquençage de l'ADN des mononucléosomes obtenus par MNase-seq dans des cellules K562. Les scores de la densité nucléosomale dans une fenêtre de ± 5 kb autour des sites d'intégration est indiqué pour les vecteurs WT et mutants. La ligne grise représente la moyenne de densité en nucléosomes à des sites choisis aléatoirement sur l'ensemble du génome.

avons dans un premier temps vérifié que le profil du vecteur sauvage correspondait à celui décrit dans la littérature (intégration favorisée dans l'euchromatine, enrichissement de modifications d'histones associées à la transcription) et nous l'avons comparé aux mutants (**Figure 41**). Les analyses montrent que le profil global des enzymes est inchangé, avec une préférence d'intégration dans les gènes transcriptionnellement actifs comparé aux régions non géniques et non transcrites. Cependant, les trois mutants R231G/A/H s'intègrent significativement plus dans les gènes que l'enzyme sauvage et le D253H (3,3 ; 6,2 et 5,9% respectivement), avec une diminution des sites d'intégration dans les régions intergéniques. Et même au sein des gènes, les mutants R231A/H sont relocalisés dans des zones plus transcrites (11,2% de plus que le WT) au détriment des zones non transcrites (entre 7 et 10% de moins). Globalement, le mutant D253H montre une tendance pour le phénotype opposé aux mutants R231, mais les résultats ne sont pas significativement différents. En résumé, les mutants R231G/A/H s'intègrent moins efficacement dans les cellules que l'enzyme sauvage, mais quand ils parviennent à s'intégrer, ils le font dans des régions différentes du WT : des régions qui sont à la fois plus riches en gènes et transcriptionnellement plus actives.

L'analyse de la densité nucléosomale de la chromatine au niveau des sites d'intégration des différents vecteurs (**Figure 42**) montre que le WT s'intègre dans des régions de la chromatine riches en nucléosomes, alors qu'au niveau du site d'intégration la densité en nucléosomes est beaucoup plus faible, ce qui est le profil attendu pour le VIH. En revanche, les mutants R231A/G/H s'intègrent dans des régions de plus faible densité en nucléosomes par rapport au WT.

L'ensemble de ces données suggère donc que l'interaction entre l'intégrase et la queue d'histone H4 influence l'efficacité d'intégration mais également le choix du site d'insertion dans la chromatine. Même si nous ne connaissons pas exactement le rôle de cette interaction au cours de l'intégration, nous proposons qu'elle puisse moduler l'association fonctionnelle entre l'intasome et le nucléosome et que dans le cas d'un mutant interagissant moins avec le nucléosome, il soit redirigé vers des zones plus pauvres en nucléosomes, régions plus ouvertes de la chromatine et transcriptionnellement très actives. Cependant, le rôle exact de l'interaction IN-queue d'histone H4 au cours de l'intégration reste à être pleinement déterminé.

7 Conclusion et discussion

Par diverses approches nous avons mis en évidence une interaction directe entre une intégrase rétrovirale et une queue d'histone. La modélisation moléculaire nous a permis d'identifier une zone dans le domaine C-terminal de l'IN où un peptide dérivé de la queue d'histone H4K20me1 pouvait se fixer. L'approche de mutagenèse dirigée nous a permis de valider l'implication de cette région dans l'interaction avec le peptide. Finalement, nous avons observé que des mutants dont l'interaction avec l'histone H4 était altérée *in vitro*, avaient

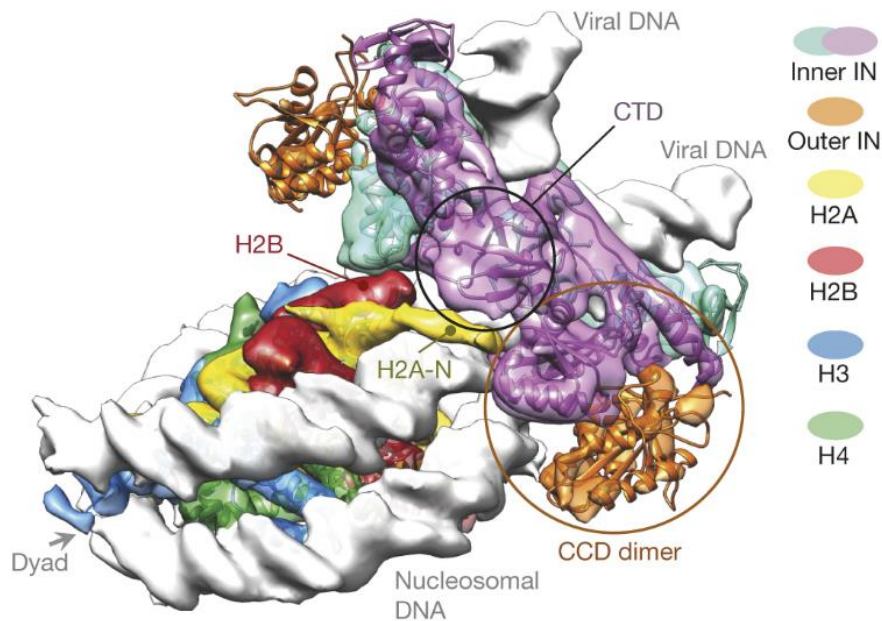


Figure 43 : Structure du complexe intasome de PFV-nucléosome. Carte de la densité électronique en surface semi-transparente avec les structures de l'intasome et du nucléosome représentées en ruban. H2B, la queue N-terminal de H2A (H2A-N) et le CTD interagissant avec cette queue sont indiqués¹²⁷.

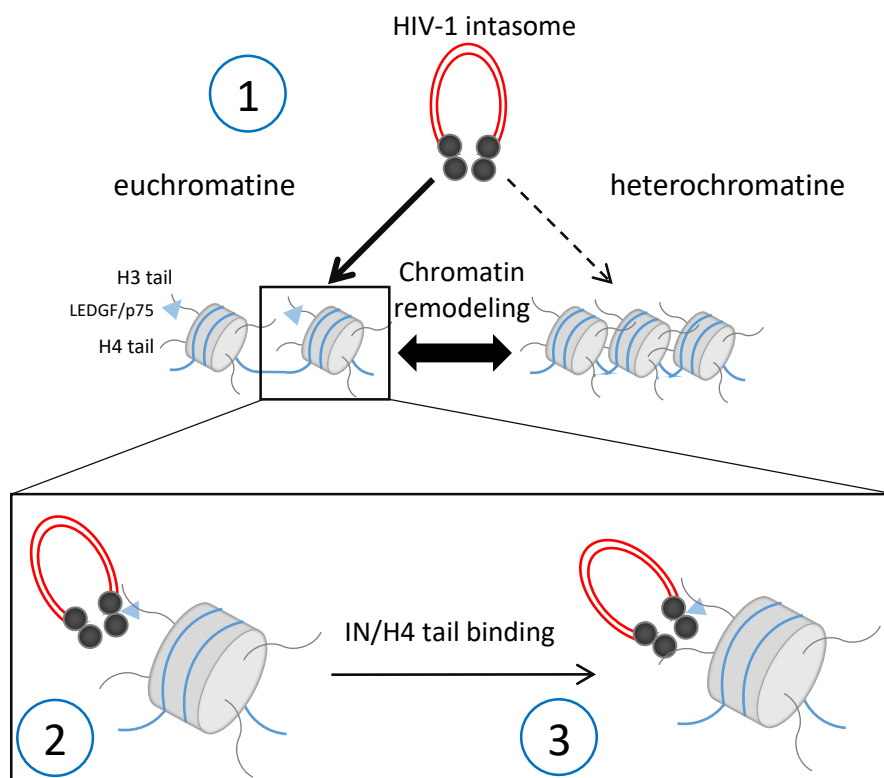


Figure 44 : Modèle d'intégration du VIH-1 faisant intervenir la queue d'histone H4. Suite à son import nucléaire, l'intasome du VIH-1 est guidé dans des zones peu denses de la chromatine grâce à CPSF6, et enrichies en LEDGF/p75 **(1)**. La queue d'histone H4 semble avoir un faible impact sur ce ciblage. L'intasome interagit ensuite avec un nucléosome via LEDGF/p75 **(2)** et requiert une association efficace avec la queue d'histone H4 **(3)** pour s'ancrer correctement sur le nucléosome afin de catalyser optimalement l'intégration.

également un défaut d'intégration *in vivo* accompagné d'un changement de sélectivité où l'intégration était redirigée dans des zones encore plus actives transcriptionnellement.

Notre hypothèse pour l'explication de ce défaut d'intégration, est que les IN mutées sont déficientes pour l'interaction avec un cofacteur d'intégration (qui est la queue d'histone H4 dans notre cas). Dans cette étude, tous nos mutants sont capables d'effectuer correctement les étapes précoces du cycle viral jusqu'à l'import nucléaire (voir quantification des différentes populations d'ADN). Nous supposons donc que le défaut d'infectivité de ces mutants est dû à un défaut pour une étape se situant après l'import nucléaire jusqu'à l'association finale de l'intasome au nucléosome. Etant donné que dans cette étude les mutants n'ont pas de défaut d'interaction avec LEDGF (non montré), nous proposons que la perte d'interaction entre l'IN et la queue d'histone H4 (et donc avec le nucléosome) est la raison pour laquelle il y a un défaut d'intégration.

Cependant la zone identifiée du CTD grâce à la modélisation moléculaire de l'interaction avec le peptide H4 a été identifiée comme impliquée dans plusieurs étapes du cycle viral, notamment le résidu R231, impliqué dans l'interaction avec la transcriptase inverse ainsi qu'avec l'ADN nucléosomal. Cette zone étant sensible à muter compte tenu de son implication pour le bon déroulement du cycle viral, il y a toujours un risque que les effets que nous avons observés en cellules soient la résultante d'un biais que nous n'avons pas identifié. Afin de compléter cette étude, d'autres mutations ont été introduites sur des résidus différents et sont en cours d'analyse au laboratoire pour valider nos hypothèses de travail.

De manière intéressante, il a été montré en parallèle de notre étude que l'intasome de PFV est également dépendant d'histones pour une association et une intégration optimale sur mononucléosome¹²⁷. La structure Cryo-EM du complexe intasome de PFV-nucléosome a révélé des contacts directs entre certains CTD de l'IN de PFV et la queue d'histone H2A (**Figure 43**). Cette interaction s'est révélée importante pour l'association de l'intasome avec le nucléosome ainsi que pour l'intégration *in vitro* et *in vivo*. De plus, des mutations altérant l'interaction de l'intasome avec le nucléosome ont également été montrées comme jouant un rôle dans la sélectivité de PFV. Dans leur étude, il semble que l'implication de l'interaction intasome-queue d'histone soit de permettre un ancrage optimal de l'intasome sur son substrat, et étant donné les phénotypes que nous avons obtenus jusqu'à présent dans notre étude, il n'est pas exclu que le rôle de la queue d'histone H4 au cours de l'intégration du VIH joue un rôle similaire au rôle de la queue d'histone H2A pour l'intégration de PFV.

Nos données *in vitro* indiquent que l'IN a une préférence nette pour la forme monométhylée de la lysine 20 (H4K20me1) comparée aux autres formes (H4K20me0, H4K20me2 ou H4K20me3). De plus, H4K20me1 est une modification épigénétique associée à la transcription active des gènes retrouvée enrichie au niveau des sites d'intégration du VIH-1¹⁵⁹. Nous n'avons pour le moment pas identifié si cette monométhylation était importante pour l'intégration *in vivo*, mais il semble que l'interaction de l'IN avec la queue d'histone H4 soit impliquée dans ce processus (**Figure 44**).

RESEARCH

Open Access



Modulation of the functional association between the HIV-1 intasome and the nucleosome by histone amino-terminal tails

Mohamed S. Benleulmi^{1,10†}, Julien Matysiak^{1,10†}, Xavier Robert^{2†}, Csaba Miskey^{3†}, Eric Mauro^{1,10}, Delphine Lapaillerie^{1,10,11}, Paul Lesbats^{1,10}, Stéphane Chaignepain⁴, Daniel R. Henriquez⁵, Christina Calmels^{1,10,11}, Oyindamola Oladosu⁶, Eloïse Thierry⁷, Oscar Leon⁵, Marc Lavigne^{8,9,11}, Marie-Line Andreola^{1,10,11}, Olivier Delelis^{7,11}, Zoltán Ivics³, Marc Ruff^{6,11}, Patrice Gouet^{2,11} and Vincent Parissi^{1,10,11*}

Abstract

Background: Stable insertion of the retroviral DNA genome into host chromatin requires the functional association between the intasome (integrase-viral DNA complex) and the nucleosome. The data from the literature suggest that direct protein–protein contacts between integrase and histones may be involved in anchoring the intasome to the nucleosome. Since histone tails are candidates for interactions with the incoming intasomes we have investigated whether they could participate in modulating the nucleosomal integration process.

Results: We show here that histone tails are required for an optimal association between HIV-1 integrase (IN) and the nucleosome for efficient integration. We also demonstrate direct interactions between IN and the amino-terminal tail of human histone H4 in vitro. Structure/function studies enabled us to identify amino acids in the carboxy-terminal domain of IN that are important for this interaction. Analysis of the nucleosome-binding properties of catalytically active mutated INs confirmed that their ability to engage the nucleosome for integration in vitro was affected. Pseudovirus particles bearing mutations that affect the IN/H4 association also showed impaired replication capacity due to altered integration and re-targeting of their insertion sites toward dynamic regions of the chromatin with lower nucleosome occupancy.

Conclusions: Collectively, our data support a functional association between HIV-1 IN and histone tails that promotes anchoring of the intasome to nucleosomes and optimal integration into chromatin.

Keywords: Retroviral integration, HIV-1, Integrase, Chromatine, Nucleosome, Histone tails

Background

Retroviral integrases (INs) are key enzymes that catalyze the insertion of viral DNA into infected cells genome (for a recent review see [1]). Integration occurs in strongly preferred regions of the genome that depend on the

virus. Although the IN is a major viral determinant in the integration site selection [2], cellular targeting factors such as BET or LEDGF/p75 proteins, which bind specific histone marks, also contribute to this process by interacting with the IN-viral DNA complex (i.e., the intasome) in these specific chromatin regions (reviewed in [3]). Additional parameters, such as the nuclear import pathway, the nuclear architecture and the interaction of cellular factors like CPSF6 with other viral components, also affect retroviral integration selectivity [4]. Thus, integration site selection is a multi-step process that first involves a global targeting of the intasome toward

*Correspondence: vincent.parissi@u-bordeaux.fr

†Mohamed S. Benleulmi, Julien Matysiak, Xavier Robert and Csaba Miskey have contributed equally to this work

¹ Fundamental Microbiology and Pathogenicity Laboratory, UMR 5234 CNRS-University of Bordeaux, SFR TransBioMed, 146 rue Léo Saignat, Bordeaux Cedex, France

Full list of author information is available at the end of the article

a suitable chromatin region via the association between IN and cellular factors, followed by local insertion step requiring IN-nucleosome interaction.

This final association between IN and its nucleosomal target substrate is a process governed by the intasome and nucleosomal DNA constraints and regulated by nucleosome density and remodeling activities [5–8]. Indeed, the data from the literature also indicate that while HIV-1 integration occurs at the surface of the nucleosomes, their compaction into dense chromatin limits efficient integration [6, 8]. We have previously shown that chromatin remodeling processes overcome this integration inhibition and favor HIV-1 integration [8]. Furthermore, we have recently reported that local nucleosome dissociation by the FACT histone chaperon generates chromatin structures favoring HIV-1 integration both in vitro and in cells [9]. Taken together these data suggest that additional contacts between the HIV-1 intasome and the nucleosome, which may be prevented during compaction and made accessible during chromatin remodeling, could be required for efficient integration. This hypothesis is supported by the cryoEM structure of the PFV intasome in complex with a mononucleosome showing direct interactions between IN protomers and histones [10]. Moreover, integration assays performed on DNA mini-circles (MCs) mimicking the nucleosomal DNA structure in the absence of histones also suggested that both this structure and additional IN/histone interactions can act in synergy during nucleosomal integration [11]. Consequently, due to the lack of information regarding the mechanisms of nucleosome capture by the HIV-1 intasome, we investigated the potential role of IN/histone interactions in regulating HIV-1 integration.

Using various biochemical and cellular approaches, we show that histone tails are required for efficient HIV-1 IN binding to nucleosomes and optimal integration. We also report that IN binds preferentially to the amino-terminal peptide tail of histone H4 (H4) in vitro and this binding is required for efficient functional interaction between the intasome and the nucleosome. Mutations affecting the IN/histone tail interaction also affect the integration step in cells. Consequently, our data lead us to conclude that the direct interaction between HIV-1 IN and histone tails may facilitate the tethering of the retroviral intasome to the nucleosomes for efficient integration into the host genome.

Results

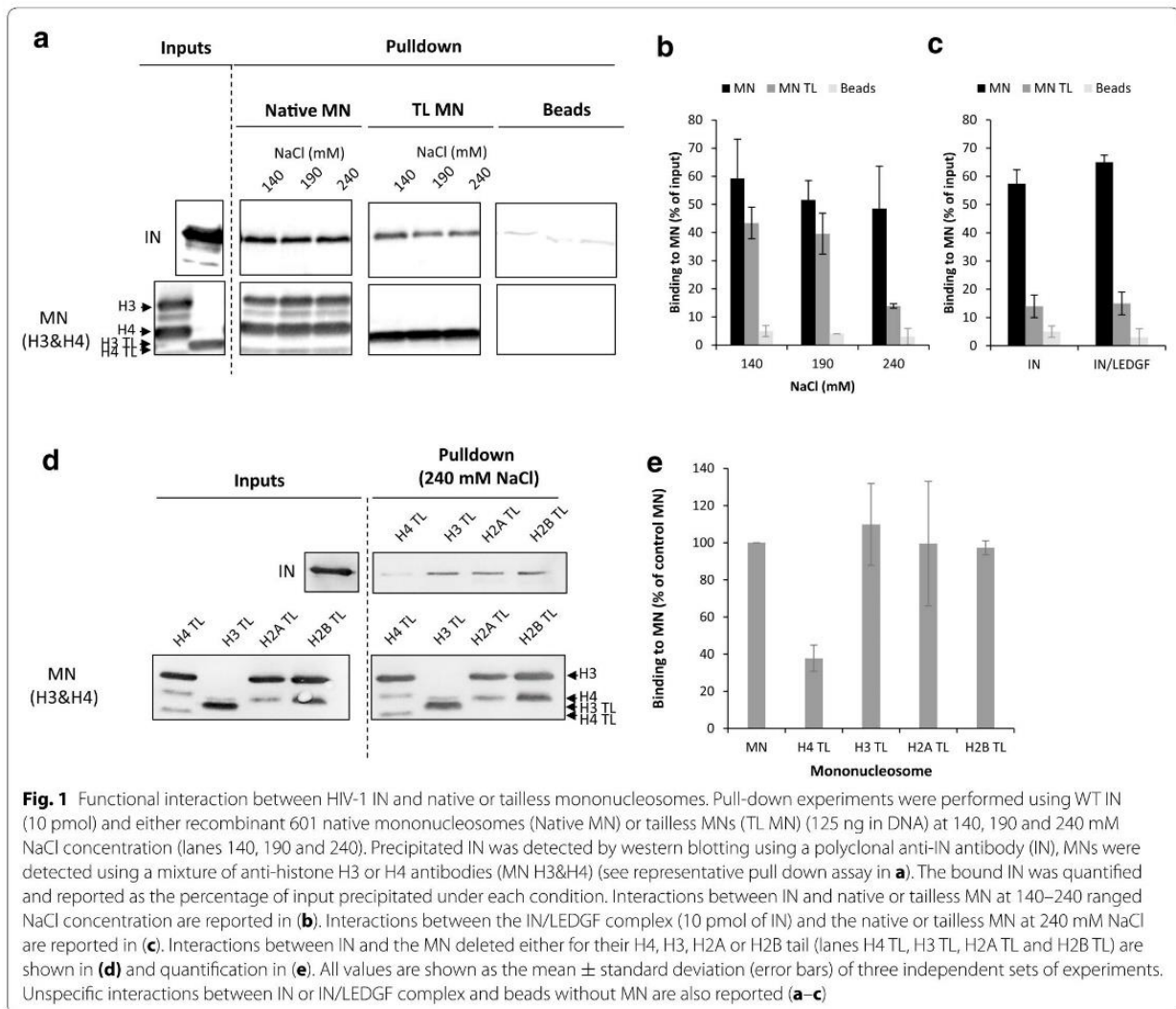
Amino-terminal histone tails modulate the interaction between HIV-1 IN and the nucleosome in vitro

To determine whether the presence of histone tails was required for the association between HIV-1 IN and the nucleosome, we performed in vitro pull-down

experiments using recombinant purified IN and either native human mononucleosomes (MNs) or tailless MNs (TL MNs) assembled on the previously described 147-bp W601 Widom sequence [12] biotinylated on its 5' end (see the MN assembly analysis in Additional file 1: Figure S1). As shown in Fig. 1, IN exhibited different affinities for native MNs and TL MNs. Indeed, increasing salt concentrations decreased the association between IN and TL MNs more efficiently than the association between IN and native MNs (Fig. 1a, b). Similar results were obtained with the IN-LEDGF/p75 complex, indicating that this functional complex also required the presence of native tails for optimal association with the nucleosome (Fig. 1c). To better determine the contribution of each histone tail in the IN/MN binding, we next performed pull-down experiments with MNs assembled using octamers lacking the tails of either H4, H3, H2A or H2B. As shown in Fig. 1d, e, the efficiency of IN binding to the H4TL MNs was approximately 50–60% less efficient than for the native and other MN variants. Interestingly, the deletion of all the histone tails had a larger impact on IN/MN binding than deletion of the H4 tail only. This may indicate that several histone tails could participate together in the binding process, the histone H4 tail appearing the most important protein determinant of this binding. To further determine the impact of histone tails on active IN/viral DNA intasomes, we next performed functional integration assays using the different MN variants.

Amino-terminal histone tails modulate the integration into nucleosomes catalyzed by HIV-1 IN in vitro

The impact of histone tails on integration activity was then evaluated in in vitro integration assays. For this purpose, the quantitative assay schematized in Fig. 2a was set up using MNs immobilized on streptavidin beads, recombinant IN and a viral DNA donor carrying the 40/42 final base pairs of the HIV-1 U5 sequence (see the "Methods" section for the description of the donor DNA). Optimized reaction conditions set up in the presence of PEG and DMSO (see materials and methods section) were first used to allow analysis of IN activity in the absence of LEDGF. The quantification of the radioactivity that remained on the beads after the reaction, washing and deproteinization, allowed us to quantify the integration efficiency. Control experiments first showed that viral DNA integrated more efficiently into MNs than into naked DNA (Fig. 2b). This result confirmed very early data reporting that MNs are the preferred substrate for HIV-1 integration [13, 14] and validated our system. Integration kinetics experiments showed that viral DNA integrated less efficiently into TL MNs than into native MNs (Fig. 2c). Speed and efficiency of integration were also



decreased when H4TL MNs were used, but to a lesser extent. Notably, integration efficiency was found to be lower when using TL MNs than when using H4TL MNs, suggesting that several histone tails could act in concert for optimal integration as suggested by the binding data. Deletion of the H3 tail slightly increased the integration efficiency, while deletion of the tails of other histone variants had no significant effect on the global integration efficiency. The presence of LEDGF/p75 did not alter the effect of histone tail deletion on integration under these conditions (Fig. 2d) and even when non-optimized reactions allowing a maximal LEDGF/p75 stimulatory effect were used (i.e. without PEG and DMSO, Additional file 1: Figure S2).

Taken together, these data indicate that native amino-terminal histone tails are required for optimal IN binding to MNs and efficient integration in vitro. Binding

experiments between IN and histone tails were next performed to further investigate whether this integration modulation could be due to such direct interactions.

Interaction between HIV-1 IN and histone amino-terminal peptide tails

Possible direct interactions between HIV-1 IN and histone tails were analyzed using a far dot blot approach with recombinant IN and peptides derived from the H3, H4, H2A and H2B amino-terminal tail (see peptide sequences in Additional file 1: Figure S3). As reported in Fig. 3a and quantification in b, interaction was significantly detected only in the presence of the histone H4 tail. Similar results were obtained with the purified IN-LEDGF/p75 complex, indicating that the LEDGF/p75 cofactor did not affect IN binding to the peptide (Fig. 3c). Additional analyzes showed that the IN/H4 tail

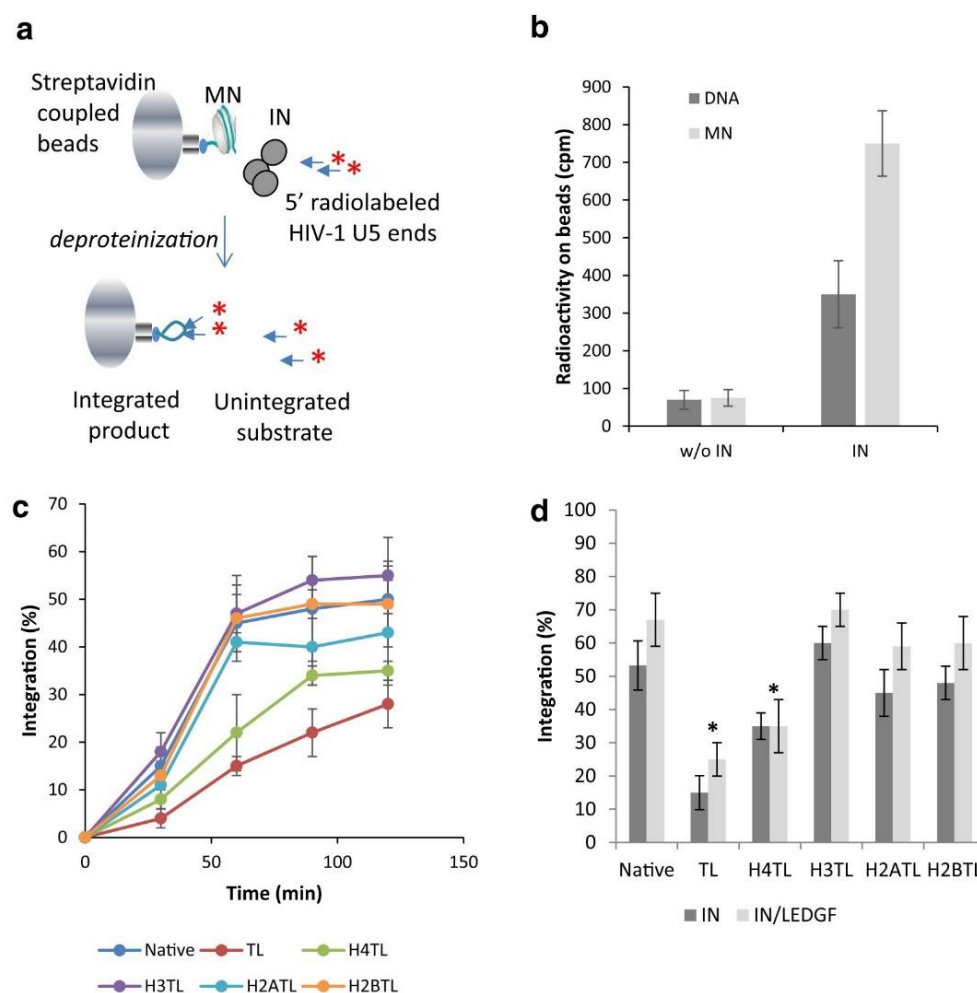


Fig. 2 In vitro Integration onto mononucleosomes. Either the 5' biotinylated naked 601 DNA fragment or the native MNs (50 ng in DNA) were coupled to streptavidin beads and incubated with HIV-1 WT IN (400 nM) under integration conditions reported in the "Methods" section (a). After 0–2 h incubations the samples were deproteinized and washed after beads magnetization, then radioactivity was measured on both the pellet and supernatant. Quantification of the radioactivity remaining on beads after reaction performed with naked 601 DNA or MN and with or without IN is reported (b). The percentage of integrated product over time for each MN construct was reported in (c). Comparison of data obtained with IN alone and IN/LEDGF complex is reported in (d). All values are shown as the mean $\pm$ standard deviation (error bars) of three to four independent sets of experiments. The p values were calculated by Student's t-test and are shown as * $p < 0.05$ to represent the probability of obtaining significant differences compared with the data obtained with the native MNs control

interaction could be negatively or positively modulated by amino acid modifications as methylation of K20 or K20 or K16 acetylation (Additional file 1: Figure S4).

The far dot blot approach was then adapted to compare different IN truncation mutants in order to identify the IN domains involved in the interaction to H4 tail. Under these conditions, the engineered IN 50–288 amino acid construct lacking the amino-terminal domain (Δ NTD) and the isolated 220–288 amino acid CTD domain construct (CTD) show similar binding properties when compared to the wild-type (WT) enzyme (Fig. 3d). By contrast, the association with the histone H4 tail was

almost completely abolished for the 1-212 amino acid construct lacking the carboxy-terminal domain (Δ CTD). These results show that the CTD domain is responsible for the interaction between IN and histone tail. In order to study the role of this interaction in the integration process we further searched for specific amino-acids mutations that could affect the IN binding to the tail.

Identification of IN mutations affecting the binding to histone H4 tail

We first adopted an in silico blind docking simulation approach starting from a fragment spanning residues

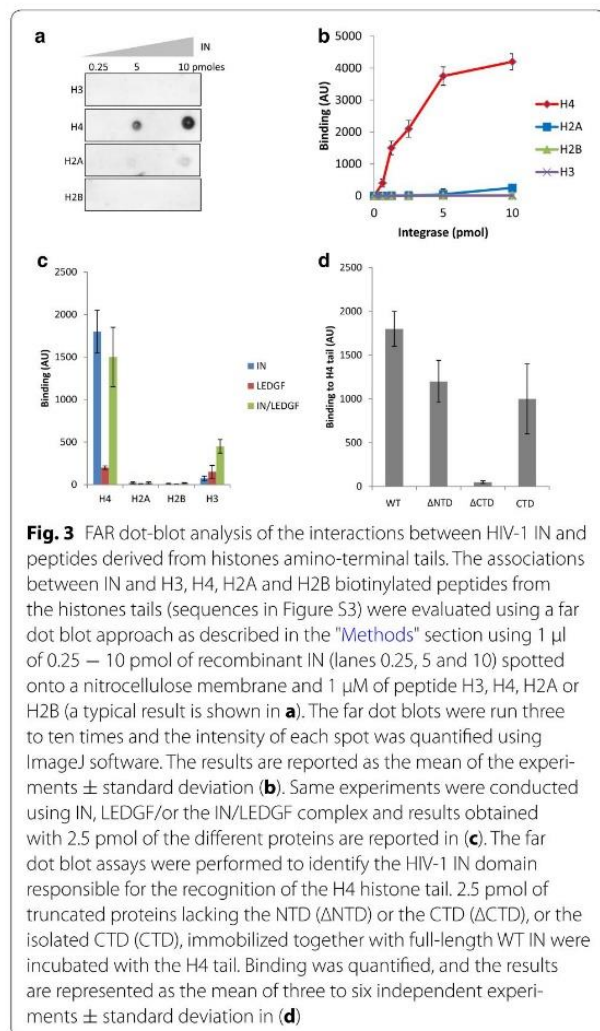


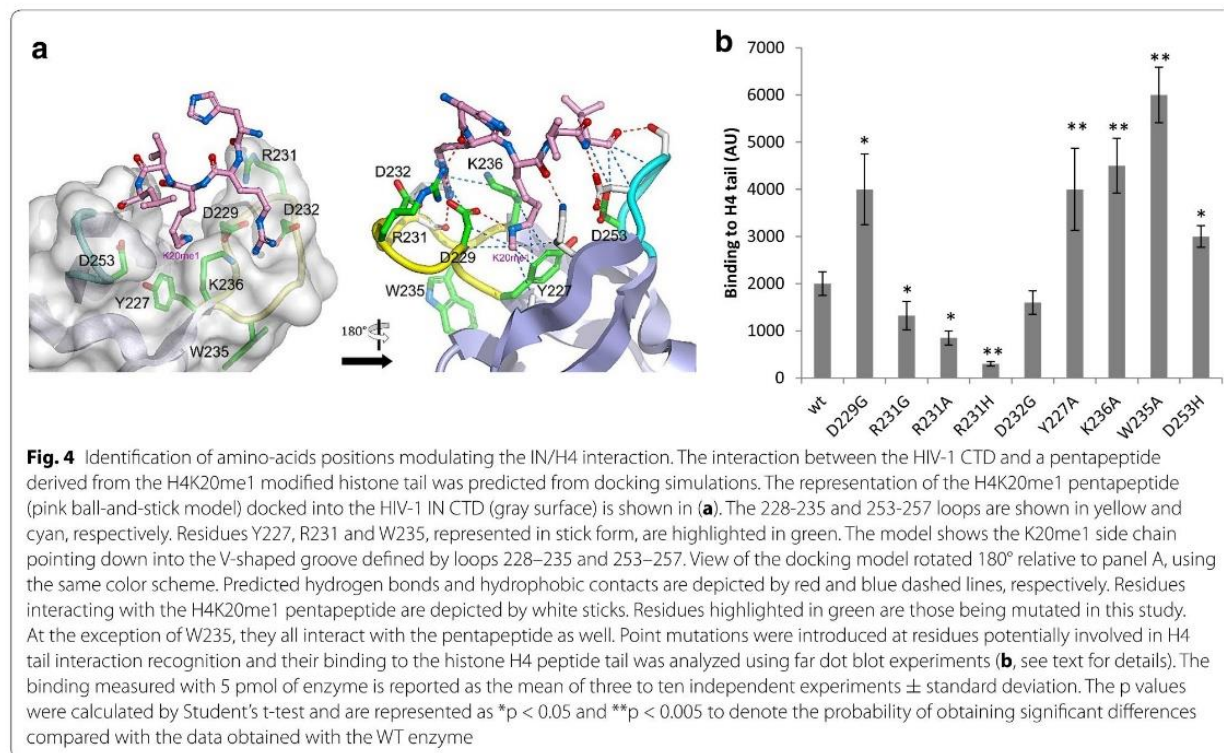
Fig. 3 FAR dot-blot analysis of the interactions between HIV-1 IN and peptides derived from histones amino-terminal tails. The associations between IN and H3, H4, H2A and H2B biotinylated peptides from the histones tails (sequences in Figure S3) were evaluated using a far dot blot approach as described in the "Methods" section using 1 μ l of 0.25 – 10 pmol of recombinant IN (lanes 0.25, 5 and 10) spotted onto a nitrocellulose membrane and 1 μ M of peptide H3, H4, H2A or H2B (a typical result is shown in **a**). The far dot blots were run three to ten times and the intensity of each spot was quantified using ImageJ software. The results are reported as the mean of the experiments $\pm$ standard deviation (**b**). Same experiments were conducted using IN, LEDGF/or the IN/LEDGF complex and results obtained with 2.5 pmol of the different proteins are reported in (**c**). The far dot blot assays were performed to identify the HIV-1 IN domain responsible for the recognition of the H4 histone tail. 2.5 pmol of truncated proteins lacking the NTD (Δ NTD) or the CTD (Δ CTD), or the isolated CTD (CTD), immobilized together with full-length WT IN were incubated with the H4 tail. Binding was quantified, and the results are represented as the mean of three to six independent experiments $\pm$ standard deviation in (**d**)

210–270 from the 2.8 Å resolution HIV-1 IN CCD-CTD structure [15] and a pentapeptide mimicking the 18–22 residues from the H4K20me1-modified histone ($H_{18}RK_{me}VL$), which corresponds to the best IN binder in the previous analyzes (see Additional file 1: Figure S4). In the first set of experiments, the AutoDock and AutoDockVina programs were used in parallel to determine a potential binding region based on a blind docking analysis of the entire surface of the receptor, namely, the IN CTD fragment, which was treated as rigid. Following a cluster analysis of all docked conformations computed by AutoDock, a potential binding site emerged in the HIV-1 IN CTD encompassing a V-shaped groove area delineated by loops 228–235 and 253–257 (one connecting strands β 1 and β 2 and the other connecting β 3 and β 4, respectively) (Fig. 4a). The resulting docking solution is compatible with the 3.9 Å resolution cryoEM structure of

the HIV-1 strand transfer complex (STC) intasome [16], in which the V-shaped CTD grooves are accessible in all the assembled IN protomers.

To determine the IN residues that may be involved in the CTD-H4 tail interaction, we focused on this latter region, where several amino acid side-chains surrounding the V-shaped groove of the receptor were treated as flexible (namely, Y227, D229, S230, R231, D232, L234, W235, K236, D253, N254, S255, D256, K258 and K264). RMSD cluster analysis of 1000 independent docking solutions using the AutoDock program allowed 56 distinct conformational groups to be defined. Considering the binding energies one solution stood out in particular, where the peptide was engaged in a total of 7 intermolecular hydrogen bonds (with the side-chains of D229, R231, S255, D256, and K258 and the backbone of L234 in the HIV-1 CTD) and 15 hydrophobic contacts (with the side-chains of Y227, D229, D232, K236, D256, K258 and V260 and the backbones of D229, S255 and D256). In this model, the peptide adopted an elongated shape at the surface of the IN CTD, with the H4K20me1 side-chain pointing down into the V-shaped groove, and formed 9 of the 16 predicted hydrophobic contacts (involving Y227, D229, K236, K258 and V260 HIV-1 CTD amino-acids residues) as well as one hydrogen bond (with D229) (Fig. 4a). Slight side-chain movements were observed to accommodate the pentapeptide, with the exception of R231 IN residue, whose side-chain flipped to form a hydrogen bond with H18 from histone 4 tail. This model was used to design a site-directed mutagenesis approach. The CTD domain has been shown to be involved in multiple functions during the viral life cycle, including interactions with reverse transcriptase and target DNA [17–19]. This made it difficult to generate CTD mutants that only affected histone binding. We focused on amino acids Y227, D229, R231, W235, K236 and D253, which were expected (1) to be located in the V-shaped groove of the IN CTD and (2) to be involved directly or indirectly in modulating the interaction. Alanine, glycine or histidine substitutions were introduced at the chosen positions to test peptide binding. The D232G substitution was also included because it represents a natural polymorphism in HIV-1 IN.

All mutants were purified, and their overall functional structures were examined in in vitro concerted integration assay. As shown in Additional file 1: Figure S5, the Y227A and W235A mutations severely affected integration (90–70% loss of activity). The K236A and D229G mutations also influenced IN catalysis, but to a lesser extent (20–40% loss). By contrast, the D232G, R231G/A/H and D253H proteins were fully active. A far dot blot assay was the used to determine the ability of the mutants to bind to and recognize the histone H4 tail. The



R231G/A/H mutants showed a decrease in their overall binding to the H4 amino-terminal tail (30, 44 and 77%, respectively; Fig. 4b). Additionally, the binding properties of the D232G mutant were virtually unaffected, whereas D229G showed a global increase in H4 tail affinity. Conversely, the Y227A, W235A, K236A and D253H mutants displayed a significant increase in affinity for the histone H4 tail.

In summary, most of the designed mutations, except the natural D232G variant, significantly affected the IN binding to the H4 tail suggesting that the corresponding amino-acids position modulate the IN/H4 interaction directly or indirectly. The identified mutants were then used to further investigate the role of the IN/H4 interaction in the association with nucleosomes.

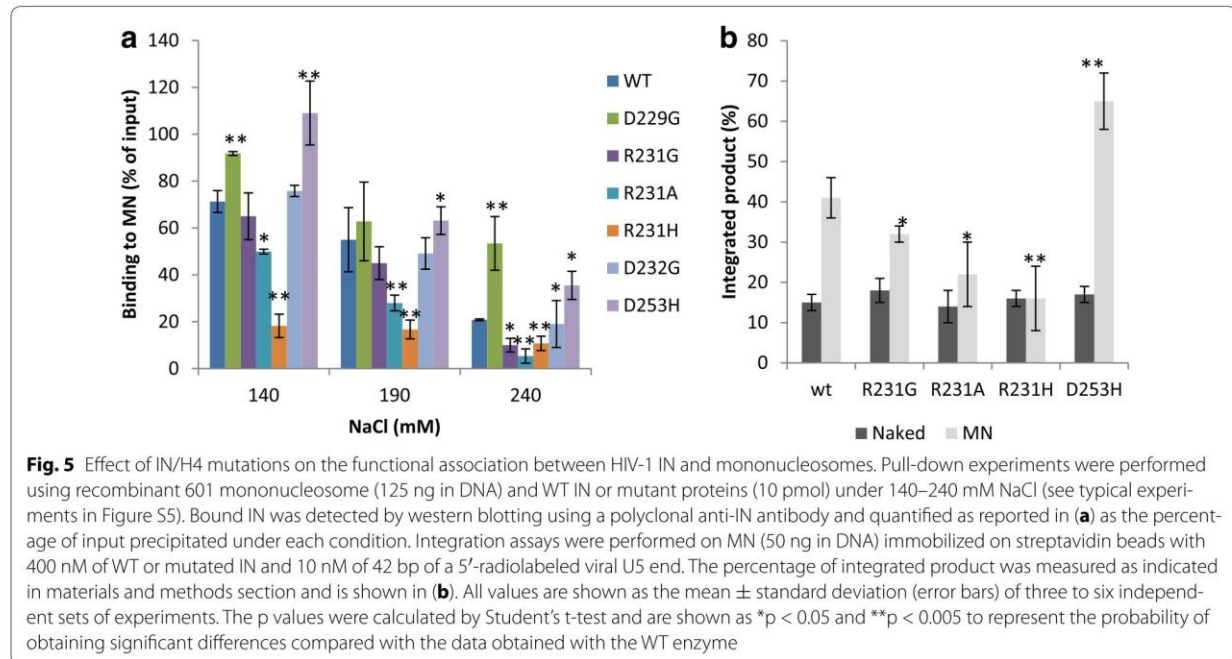
Effect of mutations affecting IN binding to H4 on the functional interaction with nucleosomes in vitro

To avoid any biases in the analysis of the MN-binding properties of the mutated INs due to the alteration of IN-DNA interaction, we first evaluated their DNA-binding properties by pull-down experiments using the naked W601 fragment. The Y227A, W235A and K236A mutants each showed decreased affinity for DNA (Additional file 1: Figure S6), which correlates well with their relative levels of in vitro integration activity. Consequently, we

excluded these enzymes from the MN interaction studies, and the mutants that showed unaffected DNA-binding capability were further analyzed for their capacity to associate with MNs.

As shown in Fig. 5a (see detailed analysis in Additional file 1: S6), the R231A/H mutants showed a significant decrease in MN binding affinity, which parallels their reduced affinity for the histone tail. The R231G mutant also had a decreased affinity for MN, but to a lesser extent, as a significant decrease in IN/MN binding was detected only at NaCl concentrations above 190 mM. By contrast, the D229G and D253H mutants, which showed an increased affinity for the H4 histone tail, also showed increased binding to MNs. The MN-binding capabilities of the natural D232G variant were not significantly affected. We next tested the effect of the mutations on the catalysis of integration into nucleosomes.

In vitro integration assays were performed using the recombinant W601 MNs used in the pull-down experiments (Fig. 5b). Control experiments performed with the unassembled W601 DNA fragment confirmed that the ability of the mutants to catalyze integration into naked DNA was not affected. In contrast, the R231G/A/H IN mutants exhibited a 25–60% decrease in efficiency of integration into MNs, and the D253H mutant was 20–40% more active than the WT enzyme. This result finely correlates with the capability of the



different INs to bind the H4 tail/MNs and fully supports our hypothesis that the binding to the tail is required for optimal integration into MNs in vitro. Therefore, we next investigated the impact of this IN/H4 interaction in a viral context.

Effect of IN/H4 mutations on viral infectivity and integration efficiency

Retroviral vectors carrying the selected R231G/A/H and D253H IN mutations, which modified the IN/H4 interaction without affecting the intrinsic IN catalytic properties, were produced, and their early replication steps were examined. The infectivity of the mutants was compared to that of WT vectors using a single-round infection assay performed in 293T cells. As shown in Fig. 6a, the infectivity of the R231G/A/H viruses was reduced by 20, 40 and 60% when compared with the WT virus, respectively. By contrast, the D253H mutation showed a 40–60% increase in viral infectivity.

The replication stages affected by the mutations were further characterized by comparing the viral DNA population size of the mutants to that of the known catalytically inactive D116A integrase (class I mutant, Fig. 6b). Under these conditions, viral cDNA production was found to be unaffected in all the viruses, indicating that there was no significant defect in the reverse transcription step, in contrast to the results observed with RT inhibition (AZT treatment). By contrast, the amount of integrated viral DNA detected for the R231G/A/H mutants was reduced by approximately 25, 60 and 80%,

respectively, with a characteristic accumulation of 2-LTR circles over time, which is indicative of normal nuclear import of the pre-integration complex. However, the D253H mutant showed a 20–40% increase in the amount of integrated DNA compared with the WT levels. This increase was associated with a decrease in the quantity of 2-LTR circles, indicating that the integration step was more efficient for this mutant, as confirmed by time-course analyses.

According to the biochemical data, one explanation for these replication phenotypes was a change in the functional association between the mutants intasomes and the chromatin/nucleosomes. To further investigate this hypothesis we next analyzed the chromatin structures surrounding the integration loci.

Effect of IN/H4 mutations on genomic integration sites selection

K562 cells were chosen because chromatin features, including histone modifications and nucleosome positions, are well annotated in this cell line. When K562 cells were transduced with lentiviruses carrying the D253H, R231G, R231A and R231H IN versions, we detected a decrease in transduction efficiency of approximately 20, 30 and 60% for the R231G/A/H mutants, respectively, and an increase in efficiency of approximately 40% for the D253H mutant compared with the WT enzyme (Fig. 7a and DNA population analyzes in Additional file 1: Figure S7). Three days post-transduction, the isolated genomic

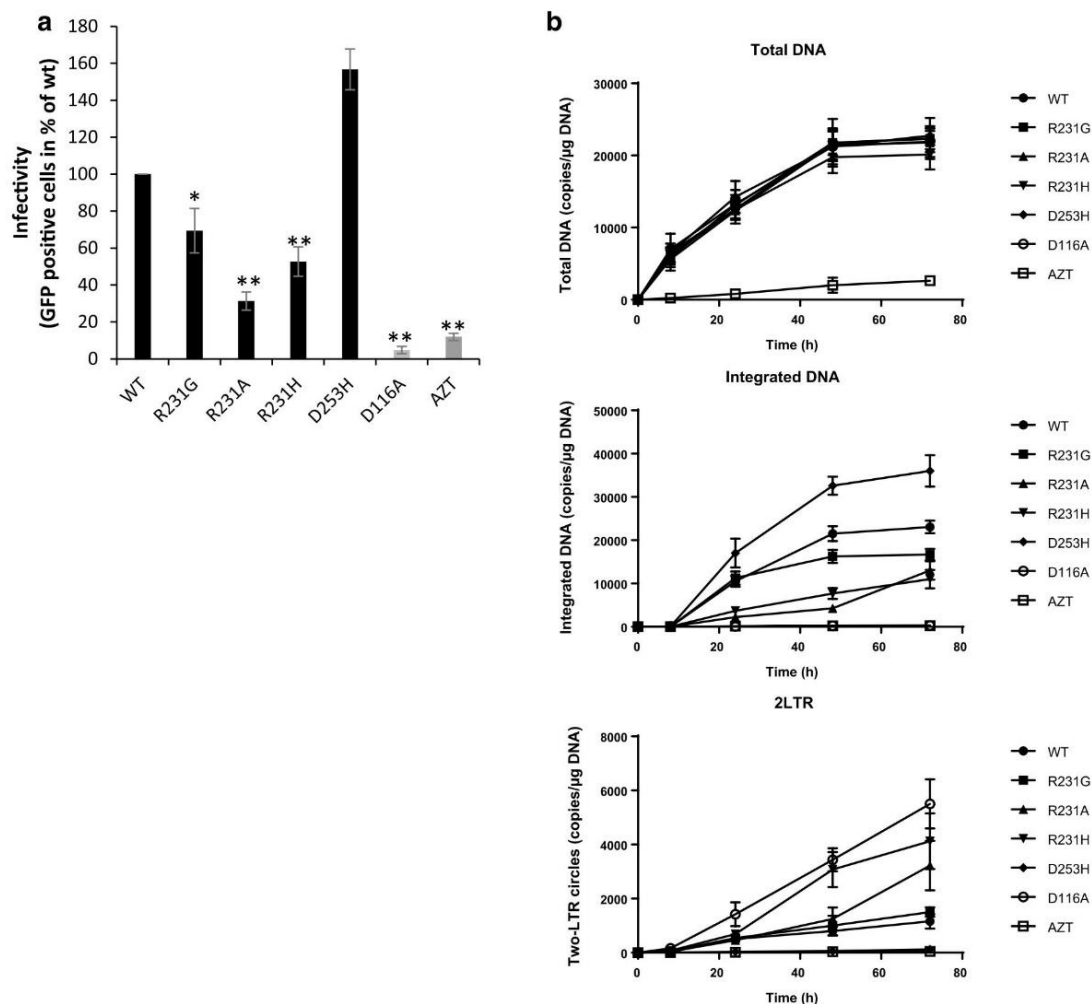


Fig. 6 Effects of mutations affecting the IN/H4 interaction during early steps of viral replication. HEK-293T cells were transduced with VSV-G pseudotyped lentiviruses encoding either WT IN or the R231A/H/G or D253H IN mutants or the catalytically inactive class I D116A mutant with or without AZT 1 μ M. Viral replication was quantified based on eGFP fluorescence measured by FACS 48 h post-transduction. The data shown in **(a)** are expressed as the percentage of eGFP-positive cells at a MOI of 1. The replication steps affected by the mutations were determined by measuring the amounts of the different viral DNA species produced using qPCR. Levels of total viral DNA, integrated DNA and 2-LTR circles shown respectively in **(b)** were monitored between 0 and 72 h post-transduction to check for potential defects at the steps of reverse transcription, integration and nuclear import of the preintegration complex, respectively. The data are represented as the mean of at least three independent experiments $\pm$ standard deviation. The p-values were calculated by Student's t-test and are shown as * $p < 0.05$ and ** $p < 0.005$ to represent the probability of obtaining significant differences compared with the WT data

DNA samples of the transduced cells were subjected to integration sites library preparation and high-throughput sequencing.

Between 4638 and 13,931 independent integration sites were obtained and analyzed. In agreement with previous findings [20, 21], analyses using genome-wide histone modification data obtained from ChIP-seq experiments performed on the chromatin of K562 cells showed that the WT insertion sites were under-represented in heterochromatin (H3K27me3-enriched

regions) and highly associated with histone marks characteristic of active transcription and open chromatin, including H3K36me3 (Additional file 1: Figure S8). We detected no significant differences in the distribution of the integration sites of the WT and the mutant INs in chromatin segments with various histone marks. By contrast, the insertion sites of the R231G/A/H mutants were more frequently localized in intragenic regions than those of the WT and D253H vectors (p value = $2.53E-4$, $3.68E-11$ and $1.68E-10$,

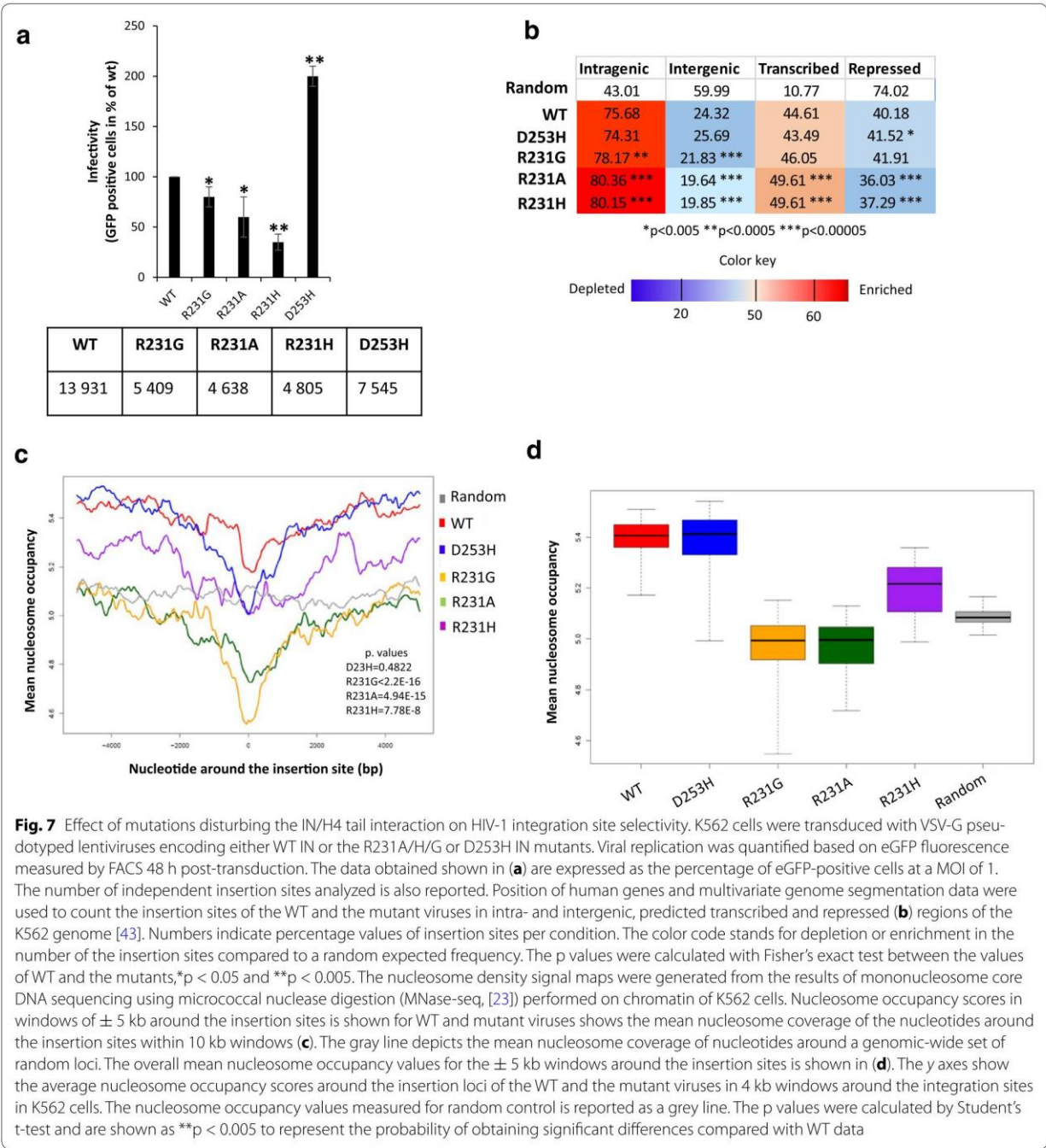


Fig. 7 Effect of mutations disturbing the IN/H4 tail interaction on HIV-1 integration site selectivity. K562 cells were transduced with VSV-G pseudotyped lentiviruses encoding either WT IN or the R231A/H/G or D253H IN mutants. Viral replication was quantified based on eGFP fluorescence measured by FACS 48 h post-transduction. The data obtained shown in **(a)** are expressed as the percentage of eGFP-positive cells at a MOI of 1. The number of independent insertion sites analyzed is also reported. Position of human genes and multivariate genome segmentation data were used to count the insertion sites of the WT and the mutant viruses in intra- and intergenic, predicted transcribed and repressed **(b)** regions of the K562 genome [43]. Numbers indicate percentage values of insertion sites per condition. The color code stands for depletion or enrichment in the number of the insertion sites compared to a random expected frequency. The p values were calculated with Fisher's exact test between the values of WT and the mutants; *p < 0.05 and **p < 0.005. The nucleosome density signal maps were generated from the results of mononucleosome core DNA sequencing using micrococcal nuclease digestion (MNase-seq, [23]) performed on chromatin of K562 cells. Nucleosome occupancy scores in windows of ± 5 kb around the insertion sites is shown for WT and mutant viruses shows the mean nucleosome coverage of the nucleotides around the insertion sites within 10 kb windows **(c)**. The gray line depicts the mean nucleosome coverage of nucleotides around a genomic-wide set of random loci. The overall mean nucleosome occupancy values for the ± 5 kb windows around the insertion sites is shown in **(d)**. The y axes show the average nucleosome occupancy scores around the insertion loci of the WT and the mutant viruses in 4 kb windows around the integration sites in K562 cells. The nucleosome occupancy values measured for random control is reported as a grey line. The p values were calculated by Student's t-test and are shown as **p < 0.005 to represent the probability of obtaining significant differences compared with WT data

respectively), and the R231 mutants integrated less frequently in intergenic territories (p value = $1.15\text{E}-5$, $3.3\text{E}-13$ and $1.6\text{E}-12$ for R231G, R231A and R231H, respectively; Fig. 7b). Additionally, the R231A/H integrase substitutions resulted in a significant increase of approximately 5% in the representations of the

integrants in transcribed regions compared with those of the WT and D253H versions (p value = $3.91\text{E}-9$ and $1.67\text{E}-8$, respectively). Concordantly, integration sites of the R231A/H mutants were less frequently found in repressed genomic territories (p value = $1.72\text{E}-20$ and $3.51\text{E}-15$, respectively). In these analyses, the R231G

mutant presented an intermediate state, as its preference for intragenic regions and transcribed genes was also affected, but to a lesser extent. Interestingly, the D253H mutant exhibited a trend opposite to that of the R231G/A/H mutants and showed a decreased preference for highly transcribed genes. In summary, we found that the R231 mutants have a stronger bias toward actively transcribed chromatin segments than the WT virus. Since the level of transcription is positively correlated with chromatin accessibility [22], we next studied the nucleosome content of the chromatin neighboring the insertion sites.

The nucleosome occupancy of the chromatin around the insertion loci was analyzed using the results of mononucleosome core DNA sequencing (MNase-seq [23]) performed on chromatin from K562 cells [22]. Similar to previous results [6, 8], measuring nucleosome occupancy in windows of ± 5 kb around the insertion sites showed that insertions of the WT vectors occurred in nucleosome-rich chromatin and that this preference declined toward the immediate insertion locus (Fig. 7c). We also found a lower mean nucleosome occupancy in the chromatin region around the R231G/A/R IN insertions sites with regards to the chromatin region surrounding the WT insertions (Wilcoxon test, $p_{R231G} < 2.2E-16$, $p_{R231A} = 4.94E-15$, $p_{R231H} = 7.78E-8$; Fig. 7c, d). These results suggest that the above vectors carrying IN/H4-disrupting mutations are less biased toward nucleosome-rich target DNA.

Since recent data suggest that residues in the HIV-1 CTD are involved in target DNA binding and recognition [7, 16, 24], we analyzed the nucleotide composition of the integration sites of the mutants. No major changes in the known weak consensus sequence of target site nucleotides typical of the WT IN were detected (Additional file 1: Figure S9). These findings, together with the results of the integration catalysis and DNA binding assays *in vitro*, argue against the possibility that the altered IN/target DNA interaction is responsible for the changes in the insertion site patterns of the mutants.

Altogether, our findings suggest that mutations disrupting the IN/H4 interaction may decrease the ability of the mutated IN to bind and functionally integrate within nucleosomes. This would explain the shift of insertion patterns toward more accessible, dynamic and nucleosome-sparse chromatin regions.

Discussion

Using multiple complementary approaches, we demonstrated that the presence of histone tails is required for efficient HIV-1 integration into nucleosomes. Additionally, we report here that HIV-1 IN binds histone amino-terminal tails, with a significant preference for the H4

tail. This interaction was shown to be required for efficient interaction with nucleosomes and optimal integration *in vitro*. Docking calculations, mutagenesis studies and binding analyses enabled us to identify several amino acid positions in the CTD of HIV-1 IN, more precisely in its V-groove, that modulate the interaction between IN and the histone tail. Analysis of the nucleosome-binding properties of the selected mutants and their capability to integrate into nucleosomes showed strong correlations between their ability to bind to the H4 tail and to nucleosomes and their ability to catalyze efficient integration into nucleosomes.

Functional analyses showed that mutations preventing the IN/H4 association also reduced viral infectivity and partly impaired the integration process. A simplest explanation for this phenotype is a deficiency in the interaction between IN and a cellular cofactor. Because all of the mutated enzymes in this study were able to interact with LEDGF/p75 (data not shown), we propose that the loss of the interaction between IN and the histone tails, leading to a loss of interaction with the nucleosome, was directly responsible for the observed integration deficiency. Importantly, the LEDGF/p75 IN cofactor did not affect IN/H4 binding or its effect on MN association and integration. This indicates that the IN/LEDGF and IN/H4 interactions may occur simultaneously, which further suggests the physiological role of this histone interaction. This is also supported by the cellular data indicating that mutations preventing the IN/H4 interaction redirect integration into genes and more dynamic regions of the chromatin. Recent studies have also reported mutations in the CTD that redirect integration. Notably, the R231G polymorphism showed more pronounced integration into GeneSeq genes but in less gene-dense and transcribed regions of the host chromatin [24]. While the redirection of this mutant into genes appears to be consistent with our data, the difference in the preference for less transcribed regions could result from differences between our and the published experimental conditions.

Interestingly, the phenotype reported in our work is reminiscent of that observed for PFV IN, which was recently reported to bind to nucleosomes via the direct interaction of IN with histones, namely, the H2A/H2B dimer surface [10]. Indeed, in both cases, PFV and HIV-1 mutants exhibiting impaired binding to MNs also showed impaired integration and an increased preference for transcribed genes and lower nucleosome occupancy regions ([10] and this work, see Fig. 7). Consequently, these data support the hypothesis that the direct binding of retroviral IN to human histones contributes to optimal integration. Retroviral intasomes may have developed various histone-binding mechanisms involving different intasome organizations.

Although several amino acid positions that modulate the HIV-1 IN/H4 interaction, including Y227, D229, R231, K236 and D253, have been identified, the putative histone-binding site has yet to be fully mapped using structural approaches. Indeed, although the mutations introduced in these positions clearly affect the association between IN and histone H4, we cannot conclude at this stage whether these positions are indirectly or directly involved in the interaction. Furthermore, the CTD has also been reported to bind target DNA [7, 33] and reverse transcriptase [17–19], making it difficult to discriminate between these pleiotropic functions and histone binding. Interestingly, the analysis of the cryoEM structure of the HIV-1 STC intasome [16] indicated that the histone tail binding site is accessible in the CTDs of all assembled IN protomers (Additional file 1: Figure S10). The CTDs of the two inner protomers contact the host DNA and are the best candidates for histone tail binding. This observation remains to be verified for the two synaptic CTDs of the lentiviral maedi-visna virus (MMV) STC intasome, whose hexadameric 4.9 Å resolution cryoEM structure reflects a plausible higher macromolecular assembly for HIV-1 IN [25]. Additionally, these recent structural data also indicate that lentiviral integration is mediated by supramolecular complexes involving a hexadecamer of IN [16, 25]. Thus, these structures show that (1) a CTD within the catalytic protomers can interact with both target DNA and the H4 tail and (2) although some CTDs of the intasome are clearly engaged with target DNA, other CTDs from other non-catalytic protomers may be available for additional protein–protein contacts. For similar reasons, it remains difficult to discriminate between the effect of R231 mutations on target DNA binding, as previously reported [7, 24], and on histone binding as reported here. However, the effect of R231 mutations on nucleotide preferences within the target site has been shown to be considerably lower than that reported for analogous PFV mutations ([16] and our own data (Additional file 1: Figure S6)). This phenotype is better explained by the recently reported structure showing a weaker interaction between the R231 HIV-1 IN residue and target DNA compared with the homologous R229 residue of PFV IN [10, 16]. This is also confirmed by the results of our integration assays and DNA binding experiments reported in Additional file 1: Figures S5 and S6 showing that the catalytic properties of these R231 mutants are not significantly affected. Furthermore, using DNA MCs mimicking the nucleosomal DNA curvature in the absence of histones, we recently showed that mutations in the CTD residues involved in target DNA binding and recognition do not significantly affect their preference for specific DNA curvatures found at the

surface of the nucleosome [11]. These data suggest that the change in target nucleosomal DNA selectivity previously observed *in vivo* [24] likely does not solely result from a loss of target DNA structure recognition but also results from a possible additional interaction with other histone-like components, as reported in our work.

Our data provide also an explanation for the inhibition of HIV-1 integration in dense chromatin templates as previously reported [6, 8]. Indeed, in these polynucleosome templates, the H4 tail is known to interact with neighboring nucleosomes, and access to the tail can be modulated by several processes, such as local chromatin remodeling [26–28]. Interestingly, the integration-refractory property of dense chromatin can be overcome by such remodeling activity [6, 8]. These data suggest that local nucleosome remodeling could be required for efficient integration by allowing additional protein/protein interactions between the incoming intasome and the nucleosome, such as the interactions between IN and histones reported herein. Moreover, we have recently shown that local remodeling by the FACT histone chaperone complex allows HIV-1 integration into poly-nucleosomes by generating partially dissociated nucleosomes which fully supports this hypothesis [9]. One direct effect of the chromatin remodeling by FACT would be thus to make accessible the H4 tails for interaction with the incoming intasomes.

Interestingly, the higher impact observed on *in vitro* integration when using tail less nucleosomes in comparison to H4 TL constructs suggests that several tails may act in synergy to modulate HIV-1 integration. Further structural determination of the intasome/nucleosome contacts by crystallography or cryo-electron microscopy, will be required to fully depict the role of each histone tails as well as histone core in the integration modulation in the context of the functional intasome/nucleosome complex.

Conclusion

The HIV-1 IN/H4 interaction reported in our work constitutes a new host/pathogen interaction important for the functional association between the incoming intasomes and the targeted nucleosome. Additional cellular processes and additional cellular protein factors, such as the recently discovered CPSF6 protein [39], participate also in regulating this multi-factorial mechanism. Consequently, optimal retroviral integration would result from an equilibrium being reached among efficient chromatin targeting, nucleosome anchoring and recognition of local DNA features. In this complex process, the interaction between IN and the H4 histone tail reported here could be an additional important determinant and, thus, constitute a potential novel therapeutic target.

Methods

Proteins, peptides and antibodies

Wild type (WT), mutated full-length and His-tagged truncated HIV-1 INs were purified as previously reported [6, 29]. GST-tagged HIV-1 IN CTD (220–288 amino-acids) was expressed in *Escherichia coli* BL21 cells (DE3) [29]. LEDGF/p75 and IN•LEDGF complex were purified as following the previously reported protocol [30, 31]. Polyclonal anti-HIV-1 IN antibodies were purchased from Bioproducts MD (Middletown, MD, USA). Antibodies directed against histones H3 (ab70550) and H4 (pAb61521 clone MABI 0400) were purchased from Abcam and Active Motif (Carlsbad, CA, USA) respectively. Recombinant mononucleosome assembled on 601 sequence biotinylated in 5' and the naked corresponding sequence were purchased from TEBU-Bio or were home-made using typical salt dialysis protocols described in [6, 8]. We used either native human histone octamers or tailless octamers purified in the Protein Expression and Purification Facility (PEPF) from the Department of Biochemistry and Molecular Biology, Colorado State University. The quality of the assembly was checked on gel shift in 0.8% agarose gel and protein content analysis on SDS-PAGE (see Additional file 1: Figure S1). Biotinylated peptides were purchased from Eurogentech (Angers, France).

In vitro integration assays

Concerted integration assays were performed as previously reported [6] using recombinant purified IN or IN•LEDGF/p75 complex (200 nM in IN monomers). IN/viral DNA complex were preassembled using previously optimized conditions [6, 32] and 10 ng of donor DNA containing the U5 viral ends (see description of the different donors in Additional file 1: Figure S11). Preassembled complexes were then incubated with 50 ng of pBSK-derived p481 plasmid DNA in 20 mM HEPES pH7, 15% DMSO, 8% PEG, 10 mM MgCl₂, 20 μM ZnCl₂, 100 mM NaCl, 10 mM DTT final concentration.

After the reaction, the resultant integration products were deproteinized by Proteinase K treatment and phenol/chloroform/isoamyl alcohol (25/24/1 v/v/v) treatment before loading onto a 1% agarose gel. The gel was then dried and submitted to autoradiography. The bands corresponding to free substrate (S), donor/donor, linear FSI (FSI) and circular HSI + FSI (HSI + FSI) products were quantified. The circular FSI products were specifically quantified by cloning them into bacteria and determining the numbers of ampicillin-, kanamycin- and tetracycline-resistant clones as percentages of the integration reaction control, which was performed using the WT enzyme. Integration assays using recombinant 601 mononucleosomes or naked 601 DNA fragments were performed using the same procedure, except that

a shorter viral DNA fragment corresponding to the 42 final base pairs of the HIV-1 U5 viral ends was used (see sequence in Additional file 1: Figure S11) and the concentration of IN was increased to 400 nM. Either 250 ng of MN or 125 ng of acceptor DNA were used. Acceptor substrates were immobilized on streptavidin-coupled beads before reaction and the reaction products were deproteinized as described above and the integration was quantified by counting the remaining radioactivity bound to magnetized beads.

Docking calculations

In all docking experiments, the fragment corresponding to residues A210–A270 from the HIV-1 IN catalytic core and the CTD crystal structure (PDB entry 1EX4) [15] was used as a protein receptor. For the ligand, we used the crystal structure of the H4K20me1 pentapeptide from the human MSL3 chromodomain complex (PDB entry 3OA6) [33]. The receptor and ligand structures were prepared for docking with AutoDockTools 1.5.6 [34]. Polar hydrogen atoms were added, non-polar hydrogens were merged, and Gasteiger partial atomic charges were computed. All possible rotatable bonds were subsequently assigned for the H4K20me1 ligand molecule. In the first set of experiments, a blind docking was performed on the entire surface of the receptor, which was treated as rigid, using the programs AutoDock 4.2.6 [34] and AutoDockVina 1.1.2 [35]. The combined docking results from these two methods enabled us to determine a unique consensus binding area. Second, experiments focusing on this area were conducted to predict the residues that may be involved in the binding of the ligand. To this end, a set of 14 residue side-chains surrounding the predicted binding area was treated as flexible. AutoGrid was used to produce grid maps that were properly centered to encompass the area of interest, with a grid box size of 76 × 84 × 98 points and a grid spacing value of 0.264 Å. AutoDock performed a total of 1000 independent runs with step sizes of 0.2 Å for translations and 5 Å for torsions. The Lamarckian Genetic Algorithm was used with a population size of 150 individuals, the maximum number of energy evaluations set to 10,000,000, the maximum number of generations set to 27,000, the maximum number of top individuals that automatically survived set to 1, and mutation and crossover rates of 0.02 and 0.8, respectively. The final cluster analysis of all docked conformations was achieved with a cluster tolerance of 3.5 Å. Finally, the top-ranked docking solutions were analyzed with AutoDockTools.

Pull-down experiments

Recombinant purified WT, mutant HIV-1 INs or IN•LEDGF/p75 complex (10 pmol of IN monomers) were

incubated with either native recombinant W601 mono-nucleosomes, tailless MNs (250 ng, i.e., 125 ng DNA), or the naked 601 DNA sequence (125 ng) in 10 μ l interaction buffer (50 mM HEPES, pH7.5; 1 μ g/ml BSA; 1 mM DTT; 0.1% Tween 20; 10% glycerol; and 50–240 mM NaCl) for 20 min on ice and then for 30 min at room temperature. A 12.5 μ l aliquot of DynabeadsMyOne Streptavidin T1 (Invitrogen, ref. 65601) was then added to a total volume of 300 μ l interaction buffer and incubated at room temperature for 1 h under rotation. The beads were washed three times with 300 μ l interaction buffer, and the precipitated products were resuspended in 10 μ l of Laemmli buffer, after which they were separated on a 12% gel via SDS-PAGE. Interacting proteins were detected by western blot analysis using anti-HIV-1 IN and anti-histone antibodies. Nucleosomal DNA was detected using a 1% agarose gel stained with SYBR[®] Safe. 140–240 mM NaCl conditions were chosen for analyzes since salt concentrations lower than 140 mM led to unspecific binding of HIV-1 IN to the beads masking its interaction with nucleosomes.

FAR dot blot experiments

One μ l of HIV-1 IN solution (1–10 pmol) was spotted onto a nitrocellulose membrane and dried for 1 h at room temperature. The membrane was then saturated for 3 h at room temperature with 5 ml of 1% BSA in PBS. After two washes, the membrane was incubated with 1 μ M of the requisite peptide in 4 ml of PBS for 1 h at 37 °C. After two washes with PBS, the membrane was incubated with ExtrAvidin coupled to horseradish peroxidase (Sigma ref. E2886 1/4000) in 4 ml of 0.3% BSA in PBS for 1 h at room temperature. The interactions were detected by ECL using a LAS4000 device. The far dot blots were run three to ten times and the intensity of each spot was quantified using ImageJ software.

Transduction of human cells with lentiviral vectors

HEK-293T (Human Embryonic Kidney 293 cells, laboratory cell line) were transduced as previously described [36]. An optimized multiplicity of infection (MOI) of 1 was used, which resulted in 25–35% of the cells containing one copy of proviral DNA as determined before. Fluorescence was quantified 48 h post-transduction by counting 10,000 cells on a FACSCalibur flow cytometer (Becton–Dickinson, San Jose, CA, USA). HIV-1 DNA species were quantified at 24, 48 and 72 h post-transduction as previously described [37]. The total and integrated HIV-1 DNA levels were determined as copy numbers per 10⁶ cells. Integrated cDNA and 2-LTR circles were expressed as a percentage of the total viral DNA.

Integration site library preparation

To remove any non-integrated viral DNA (and one-, or two-LTR circles) per condition, 5 μ g genomic DNA (gDNA) samples isolated from K562 (human immortalized myelogenous leukemia cell line purchased from ATCC company) 72 h post-transfection were subjected to 0.6% agarose gel electrophoresis and high-molecular gDNA was isolated from the gel using the Zymoclean[™] Large Fragment DNA Recovery Kit, (Zymo Research). The eluents were sonicated to an average of 600 bp-long fragments in screw-cap cuvettes with the Covaris M220 ultrasonicator with the following settings: peak power: 50.0, duty factor: 20, cycle/burst: 200, duration: 28 s. After bead purification the DNA was end-repaired and 5'-phosphorylated with the NEBNext End Repair Module (New England Biolabs, (NEB)). The DNA was prepared for ligation with NEBNextdA-Tailing Module, (NEB) and eluted after bead purification in 10 μ l water. Ligation with double-stranded linkers (see Additional file 1: Figure S10) was performed in 15 μ l for 15 min at room temperature using the Blunt/TA Ligase Master Mix (NEB). After purification with 0.8 volumes of AMPure XP beads (Beckman Coulter), the ligated DNA was eluted in 20 μ l of 10 mM Tris/HCl, pH 8.0 and the whole DNA solution was used for multiple PCR reactions to amplify the virus-gDNA junctions with the primers SIN-HIV1 and linker primer using NEBNext High-Fidelity 2 $\times$ PCR Master Mix (NEB) with the following cycling conditions: 98 °C 30 s; 20 cycles of: 98 °C 10 s, 68 °C 30 s, 72 °C 30 s; 10 cycles of: 98 °C 10 s, ramp to 63 °C 1 °C/s 30 s, 72 °C 30 s; 72 °C 3 min. The PCR products were isolated using 1 volume of AMPure XP beads (Beckman Coulter), eluted in 20 μ l of 10 mM Tris/HCl, pH 8.0 and 2 μ l of the eluents served as template for 5 parallel PCR reactions with the primers: SIN-HIV-BC-N-III and PE-nest ind-N (where N stands for the sequences of Illumina TrueSeq indexes, or their corresponding reverse complement sequences) using the following cycling conditions: 98 °C 30 s; 20 cycles of: 98 °C 10 s, 67 °C 30 s, 72 °C 30 s; 72 °C 3 min. The 200–500 bp size range of the indexed libraries were agarose gel-isolated and mixed equimolarly for 100 base, single-end Illumina sequencing on a HiSeq 2000 instrument using 40% PhiX DNA spike-in at Genewiz, USA.

Analysis of sequencing data

The raw reads starting with condition-specific indexes were grouped and filtered for the presence of the virus-specific nested primer followed by LTR sequences at the tip of the LTR. The rest of the reads were quality-trimmed as soon as 2 out of 5 bases had quality scores less than a Phred score of 20. We used *bowtie* [38] with

the TAPDANCE tool [39] to map the reads to the hg19 human genome assembly in cycles with decreasing read length of 60, 55, 50, 45, 40, 35 allowing 3, 3, 3, 2, 1, 0 mismatches, respectively, with the following bowtie parameters in the mapping cycles: `[-quiet -a -v < nu. mismatches allowed > -m 1 -suppress 5,6,7 -f]`. Any insertion site was considered valid if there were at least 5 independent reads supporting it. All read pre-processing and follow-up analyses were done in R (R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>).

Analysis of insertion sites in chromatin features

Nucleosome occupancy signal datasets for K562 cells were obtained from ENCODE [22]. Genomic coordinates with an associated nucleosome occupancy density signal value greater than zero were used to calculate occupancy matrixes and to plot nucleosome densities with the *genomation* R package [40]. BEDTools [41] and *genomation* were used to analyze the representation of ISs in histone mark distributions [42] and in chromatin state segment datasets making use of a consensus merge of the segmentations produced by the ChromHMM and Segway software [43]. We applied the Wilcoxon test on the row-sums of the score matrixes generated from nucleosome occupancy datasets to check for any statistical difference between the conditions. Fisher's exact test was used to calculate statistical significance between the representations of ISs of the WT and the integrase mutant viruses within methylated histone ChIP-seq peaks.

Additional files

Additional file 1: Figure S1. Structure of the native and tailless mononucleosomes used in the work. The globular structure of the nucleosomes was analyzed by loading 250 ng of native or tailless MN on 0.8% native agarose gel run 4 h at 50 V and 4 °C then stained with SYBR[®]Safe 20 min. Assembled MNs migrate between 600 and 700 bp and the naked 601 DNA fragment at 147 bp. **Figure S2.** A. Effect of LEDGF/p75 on HIV-1 integration in vitro. Integration assay was performed as done in Fig. 2 using naked 601 DNA coupled to magnetic beads and increasing concentration of LEDGF in the presence or absence of PEG and DMSO. B. Integration activity catalyzed by IN and IN/LEDGF complex on native or tail less nucleosomes in the absence of PEG and DMSO. Integration assay was performed as done in Fig. 2 using native or tail less nucleosomes coupled to magnetic beads and either IN or IN/LEDGF complex. All values are shown as the mean $\pm$ standard deviation (error bars) of three independent sets of experiments. The p values were calculated by Student's t test and are shown as *p < 0.05 to represent the probability of obtaining significant differences compared with the data obtained with the native MNs control. **Figure S3.** Sequence of the peptide tails used in the work. **Figure S4.** FAR dot-blot analysis of the interactions between HIV-1 IN and peptides derived from histone 4 amino-terminal tails. The associations between IN and unmodified H4, or modified H4 peptides were evaluated using a far dot blot approach as described in the "Methods" section using 1 μ l of 2.5 pmol of recombinant IN spotted

onto a nitrocellulose membrane and 1 μ M of peptides. The far dot blots were run three to ten times and the intensity of each spot was quantified using ImageJ software. The results are reported as the mean of the experiments $\pm$ standard deviation. **Figure S5.** In vitro integration activities of wild type and mutant integrases. A concerted integration assay was performed using 200 nM of different enzymes which were purified using a similar procedure, in addition to 10 ng of donor DNA and 50 ng of pBSK-derived p481 plasmid DNA. The reaction products were loaded onto 1% agarose gels and a representative set of experiments is shown in (A). The positions and structures of the donor substrate and the different half-site (HSI), full-site (FSI) and donor/donor integration (d/d) products are shown. Quantification of the total integration is shown in (B) as a percentage of WT activity. The circular FSI products were quantified by cloning them into bacteria and are shown in (C) as the numbers of ampicillin-, kanamycin- and tetracycline-resistant clones as percentages of the integration reaction control performed using the WT enzyme. All values are shown as the mean $\pm$ standard deviation (error bars) of at least three independent sets of experiments. The p-values were calculated by Student's t-test and are shown as *p < 0.05 and **p < 0.005 to represent the probability of obtaining significant differences compared with WT data set at 100%. **Figure S6.** HIV-1 IN and mononucleosome pull-down experiment. Naked 147 bp 601 DNA sequence or MN assembled on this fragment were used (structure of the naked and assembled 601 DNA is reported in the gel shift experiment shown in **Figure S1**). WT IN was efficiently pulled down using a biotinylated naked 601 DNA fragment (left panel) or 601 mononucleosomes assembled on the same DNA (right panel) immobilized on streptavidin beads using 140–240 mM of NaCl (A). Experiments were performed using different mutated enzymes. Each pull-down was run three to six times and the intensity of each band was quantified using ImageJ software. The results obtained with naked DNA are reported as the mean of the experiments $\pm$ standard deviation in (B). The p values were calculated by Student's t-test and are shown as *p < 0.05 and **p < 0.005 to represent the probability of obtaining significant differences compared with the WT data in each condition. **Figure S7.** Time course analysis of the early steps of replication of wild type and mutants viral vectors in K562 cells. K562 cells were transduced with VSV-G pseudotyped lentiviruses encoding either WT IN or the R231A/H/G or D253H IN mutants. The replication steps affected by the mutations were determined by measuring the amounts of the different viral DNA species produced using qPCR. Levels of total viral DNA, integrated DNA and 2-LTR were monitored between 0 and 72 h post-transduction to check for potential defects at the steps of reverse transcription, integration and nuclear import of the preintegration complex. The data are represented as the mean of at least three independent experiments $\pm$ standard deviation. The p-values were calculated by Student's t-test and are shown as *p < 0.05 and **p < 0.005 to represent the probability of obtaining significant differences compared with the WT data. **Figure S8.** Effect of mutations affecting the IN/H4 tail interaction on HIV-1 integration site selectivity. Integration sites of the WT and the mutant viruses were annotated in signal peaks of ChIP-seq experiments for genome-wide histone modifications in K562 cells. Numbers indicate percentage values of insertion sites per condition. The p values were calculated with Fisher's exact test between the values of WT and the mutants, *p < 0.05 and **p < 0.005. **Figure S9.** Consensus sequences directly neighboring insertion sites of pseudoviral vectors carrying IN/H4 mutations. The target DNA consensus diagrams were generated with the *seqLogo* package in R. The triangles show the insertion sites. The relative height of individual bases at each position is proportional to the frequency of the base at that position. **Figure S10.** Superimposition of the HIV-1 IN CTD-H4K20me1 docking model with the structure of the tetrameric HIV-1 strand transfer complex intasome (PDB entry 5U1C). The model is presented in magenta cartoon representation with the docked H4K20me1 pentapeptide highlighted in green. The CTDs of the two inner protomers contacting the host DNA (colored in gold) are depicted in salmon and cyan. The grey cartoon corresponds to the rest of the HIV-1 strand transfer complex structure. **Figure S11.** Sequence of the viral DNA donors used in concerted integration assays. For concerted integration on MNs the two HIV1_U5 (+) and HIV1_U5 (-) (A) were hybridized and the resulting 42/40 bp hybrid was radiolabeled in 5' with T4 DNA kinase. For concerted integration on naked DNA plasmid we used the 246 bp DNA

fragment shown in (B) generated by PCR on a pUC19 supF. After purification the fragment was radiolabeled in 5' with T4 DNA kinase. Sequence of oligonucleotides used for integration selectivity analyses (C).

Additional file 2. List of insertion sites of all the tested conditions. Raw sequencing reads are available upon request.

Authors' contributions

MSB, JM, EM, DL, PL, DRH and VP performed the in vitro assays. CC purified the full length and truncated HIV-1 IN.1 ET and OD perform the viral DNA quantification. OO and MR performed the thermophoresis experiments and purified the HIV-1 IN CTD. XR and PG performed the docking calculation. CM and ZI performed the integration selectivity analyses. MSB, XR, CM, PL, SC, OL, ML, MLA, OD, ZI, MR, PG and VP analyzed and discussed the data. MSM, XR, MR, OD, PG and VP wrote the manuscript. All authors read and approved the final manuscript.

Author details

¹ Fundamental Microbiology and Pathogenicity Laboratory, UMR 5234 CNRS-University of Bordeaux, SFR TransBioMed, 146 rue Léo Saignat, Bordeaux Cedex, France. ² MMSB-Institute of the Biology and Chemistry of Proteins, UMR 5086 CNRS-Lyon 1 University, Lyon, France. ³ Division of Medical Biotechnology, Paul Ehrlich Institute, Langen, Germany. ⁴ UMR CNRS 5248 CBMN (Chimie Biologie des Membranes et Nanoobjets), Université de Bordeaux, 33076 Bordeaux, France. ⁵ Virology Program, ICBM, Faculty of Medicine, University of Chile, Santiago of Chile, Chile. ⁶ Département de Biologie Structurale Intégrative, UDS, U596 INSERM, UMR7104 CNRS, IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), Illkirch, France. ⁷ LBPA, UMR8113, CNRS, ENS-Cachan, 94235 Cachan, France. ⁸ Dpt de Virologie, UMR 3569, CNRS, Institut Pasteur, Paris, France. ⁹ Institut Cochin-Inserm U1016-CNRS UMR8104-Université Paris Descartes, Paris, France. ¹⁰ International Associated Laboratory (LIA) of Microbiology and Immunology, CNRS, University de Bordeaux/Heinrich Pette Institute-Leibniz Institute for Experimental Virology, Bordeaux, France. ¹¹ Viral DNA Integration and Chromatin Dynamics Network (DyNAVIR), Bordeaux, France.

Acknowledgements

The authors are deeply grateful to Dr. Simon Litvak for fruitful discussions. The manuscript was edited by NPG Language Editing and Prof Ray Cooke.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files. A list of lentiviral vectors insertion sites of all the tested conditions is provided in Additional file 2: Figure S12 and raw sequencing reads are available upon request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by the French National Research Agency [ANR, RETROSelect program]; the French National Research Agency against AIDS (ANRS, AO 2016-2, ECTZ18624); SIDACTION (AO-27-1 10465, 16-1-AEQ-10465); the French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-01]; Instruct, a part of the European Strategy Forum on Research Infrastructures (ESFRI); the Centre National de la Recherche Scientifique (CNRS); the University Victor Segalen Bordeaux 2; and the ECOS-CONICYT C12B03 program.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 16 October 2017 Accepted: 20 November 2017

Published online: 28 November 2017

References

- Lesbats P, Engelman AN, Cherepanov P. Retroviral DNA integration. *Chem Rev*. 2016;116(20):12730–57.
- Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, Crawford G, et al. Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog*. 2006;2:e60.
- Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A. Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res*. 2014;42(16):10209–25.
- Sowd GA, Serrao E, Wang H, Fadel HJ, Poeschla EM, et al. A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc Natl Acad Sci USA*. 2016;113:E1054–63.
- Naughtin M, Haftek-Terreau Z, Xavier J, Meyer S, Silvain M, Jaszczyszyn Y, et al. DNA physical properties and nucleosome positions are major determinants of HIV-1 integrase selectivity. *PLoS ONE*. 2015;10:e0129427.
- Benleulmi MS, Matysiak J, Henriquez DR, Vaillant C, Lesbats P, Calmels C, et al. Intasome architecture and chromatin density modulate retroviral integration into nucleosome. *Retrovirology*. 2015;12:13.
- Serrao E, Krishnan L, Shun MC, Li X, Cherepanov P, Engelman A, et al. Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: implications for the mechanism of target DNA binding. *Nucleic Acids Res*. 2014;42:5164–76.
- Lesbats P, Botbol Y, Chevereau G, Vaillant C, Calmels C, Arneodo A, et al. Functional coupling between HIV-1 integrase and the SWI/SNF chromatin remodeling complex for efficient in vitro integration into stable nucleosomes. *PLoS Pathog*. 2011;7:e1001280.
- Matysiak J, Lesbats P, Mauro E, Lapallierie D, Dupuy J-W, Lopez AP, et al. Modulation of chromatin structure by the FACT histone chaperone complex regulates HIV-1 integration. *Retrovirology*. 2017;14(1):39.
- Maskell DP, Renault L, Serrao E, Lesbats P, Matadeen R, Hare S, et al. Structural basis for retroviral integration into nucleosomes. *Nature*. 2015;523:366.
- Pasi M, Mornico D, Volant S, Juchet A, Batisse J, Bouchier C, et al. DNA minicircles clarify the specific role of DNA structure on retroviral integration. *Nucleic Acids Res*. 2016;44:7830.
- Lowary PT, Widom J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol*. 1998;276:19–42.
- Pryciak PM, Sil A, Varmus HE. Retroviral integration into minichromosomes in vitro. *EMBO J*. 1992;11:291–303.
- Pryciak PM, Müller H-P, Varmus HE. Simian virus 40 minichromosomes as targets for retroviral integration *in vivo*. *Proc Natl Acad Sci USA*. 1992;89:9237–41.
- Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, Leavitt AD, et al. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc Natl Acad Sci USA*. 2000;97:8233–8.
- Passos DO, Li M, Yang R, Rebersburg SV, Ghirlando R, Jeon Y, et al. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science*. 2017;355:89–92.
- Lu R, Ghory HZ, Engelman A. Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase. *J Virol*. 2005;79:10356–68.
- Lu R, Limón A, Ghory HZ, Engelman A. Genetic analyses of DNA-binding mutants in the catalytic core domain of human immunodeficiency virus type 1 integrase. *J Virol*. 2005;79:2493–505.
- Tekeste SS, Wilkinson TA, Weiner EM, Xu X, Miller JT, Le Grice SJ, et al. Interaction between reverse transcriptase and integrase is required for reverse transcription during HIV-1 replication. *J Virol*. 2015;89:12058–69.
- Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, et al. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol*. 2004;2:E234.

21. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* 2007;17:1186–94.
22. Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, et al. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat Commun.* 2016;7:11485.
23. Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. Determinants of nucleosome organization in primary human cells. *Nature.* 2011;474:516–20.
24. Demeulemeester J, Vets S, Schrijvers R, Madlala P, De Maeyer M, De Rijck J, et al. HIV-1 integrase variants retarget viral integration and are associated with disease progression in a chronic infection cohort. *Cell Host Microbe.* 2014;16:651–62.
25. Ballandras-Colas A, Maskell DP, Serrao E, Locke J, Swuec P, Jónsson SR, et al. A supramolecular assembly mediates lentiviral DNA integration. *Science.* 2017;355:93–5.
26. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature.* 1997;389:251–60.
27. Dorigo B, Schalch T, Bystrycki K, Richmond TJ. Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J Mol Biol.* 2003;327:85–96.
28. Song F, Chen P, Sun D, Wang M, Dong L, Liang D, et al. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science.* 2014;344:376–80.
29. Busso D, Delagoutte-Busso B, Moras D. Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal Biochem.* 2005;343:313–21.
30. Botbol Y, Raghavendra NK, Rahman S, Engelman A, Lavigne M. Chromatinized templates reveal the requirement for the LEDGF/p75 PWWP domain during HIV-1 integration in vitro. *Nucleic Acids Res.* 2008;36:1237–46.
31. Levy N, Eiler S, Pradeau-Aubretton K, Maillot B, Stricher F, Ruff M. Production of unstable proteins through the formation of stable core complexes. *Nat Commun [Internet].* 2016 [cited 2017 Feb 16];7. <https://www.ncbi.nlm.nih.gov/insb.bib.cnrs.fr/pmc/articles/PMC4800440/>.
32. Lesbats P, Metifiot M, Calmels C, Baranova S, Nevinsky G, Andreola ML, et al. In vitro initial attachment of HIV-1 integrase to viral ends: control of the DNA specific interaction by the oligomerization state. *Nucleic Acids Res.* 2008;36:7043–58.
33. Kim D, Blus BJ, Chandra V, Huang P, Rastinejad F, Khorasanizadeh S. Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nat Struct Mol Biol.* 2010;17:1027–9.
34. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem.* 2009;30:2785–91.
35. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multi-threading. *J Comput Chem.* 2010;31:455–61.
36. Cosneffroy O, Tocco A, Lesbats P, Thierry S, Calmels C, Wiktorowicz T, et al. Stimulation of the human RAD51 nucleofilament restricts HIV-1 integration in vitro and in infected cells. *J Virol.* 2012;86:513–26.
37. Munir S, Thierry S, Subra F, Deprez E, Delelis O. Quantitative analysis of the time-course of viral DNA forms during the HIV-1 life cycle. *Retrovirology.* 2013;10:87.
38. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25.
39. Sarver AL, Erdman J, Starr T, Largaespada DA, Silverstein KAT. TAPDANCE: an automated tool to identify and annotate transposon insertion CISs and associations between CISs from next generation sequence data. *BMC Bioinform.* 2012;13:154.
40. Akalin A, Franke V, Vlahovićek K, Mason CE, Schübeler D. genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics.* 2015;31:1127–9.
41. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.
42. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, et al. Systematic analysis of chromatin state dynamics in nine human cell types. *Nature.* 2011;473:43–9.
43. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods.* 2012;9:473–6.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



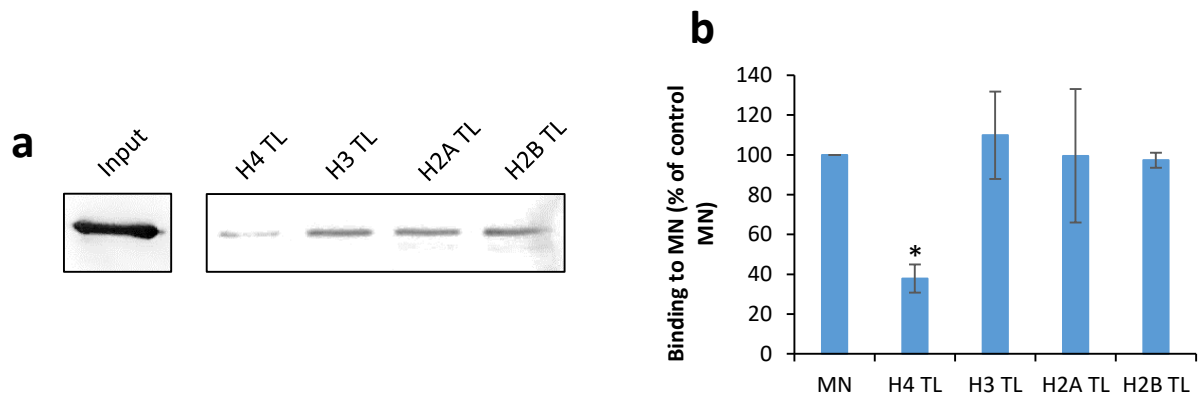


Figure 45 : Influence des queues d'histones sur l'interaction IN-MN *in vitro*. (a) Interaction entre l'IN et les différentes constructions de MN (TL = tail less) biotinylés par pull down à 240mM NaCl et la quantification de l'interaction en (b). Les écarts types ont été calculés avec un test t de Student et représentés par * $p < 0,05$ sur au moins trois expériences indépendantes.

Partie 2 : Caractérisation biochimique de l'interaction IN-H4 au cours de l'intégration

Ce travail a mené à un article en premier auteur qui est en cours de soumission : « **Human H4 aminoterminal tail stimulates HIV-1 integration through structural changes in the carboxyterminal domain of integrase** »

Mauro et al.

1 Contexte scientifique et objectifs

Grâce à l'utilisation de vecteurs lentiviraux exprimant une intégrase altérée pour l'interaction avec la queue d'histone H4, nous avons pu mettre en évidence l'importance de cette interaction pour que l'intégration du VIH-1 se fasse de manière optimale dans les cellules (voir partie précédente). Cette interaction semble influencer à la fois sur l'efficacité et la sélectivité de l'intégration. Bien que l'effet sur la sélectivité soit modéré, le mutant R231H présente une infectivité réduite d'environ 60% dans notre système, soulignant ainsi l'importance de cette interaction pour une intégration optimale. De plus, l'interaction de ce mutant avec la queue d'histone n'est pas complètement abolie, ce qui suggère que l'effet observé sur l'inhibition de l'intégration dans les cellules peut encore être amplifié, par l'identification de mutations réduisant encore davantage l'interaction avec la queue d'histone H4 ou même l'utilisation de molécules bloquant l'interaction de ces deux partenaires (ce point constitue le dernier axe du projet). Mais pour développer cette stratégie, nous devons d'abord mieux caractériser le rôle de cette interaction au cours de l'intégration.

2 Effet des queues d'histones sur l'ancrage et l'intégration du VIH sur MN

Afin de corrélérer les résultats obtenus par Maskell et al. pour PFV avec nos observations pour VIH, nous avons cherché à savoir si la queue d'histone H4 était importante pour l'ancrage de l'IN au nucléosome. Pour cela, nous avons réalisé des pull down avec de l'IN recombinante et différentes constructions de mononucléosomes biotinylés assemblées au laboratoire auxquels il manque une des queues d'histones. Comme on peut le voir en **Figure 45**, l'IN n'a pas la même affinité pour ces différentes constructions, en particulier pour le

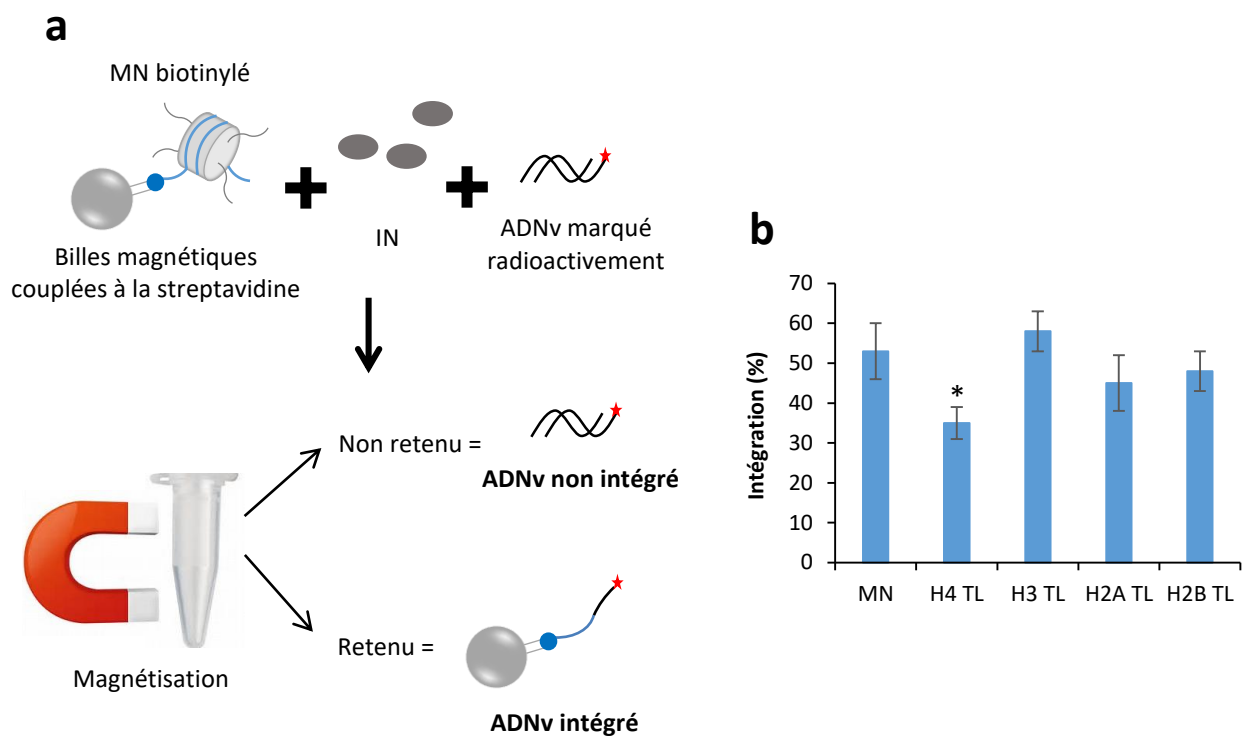


Figure 46 : Influence des queues d'histones sur l'intégration *in vitro* sur MN. (a) Schéma du test d'intégration utilisé pour quantifier l'influence des queues d'histones sur l'intégration. Les MN TL (tailless) assemblés avec la séquence d'ADN 601 biotinylée est incubé avec un ADN donneur court HU5. A l'issue de la réaction les produits intégrés et non intégrés sont séparés par magnétisation et la radioactivité est quantifiée. **(b)** Quantification des tests d'intégration réalisés sur les différentes constructions de MN. Les écarts types ont été calculés avec un test t de Student et représentés par * $p < 0,05$ sur au moins trois expériences indépendantes.

mononucléosome délété de la queue d'histone H4 pour laquelle l'IN perd plus de 60% d'interaction dans ces conditions par rapport à un mononucléosome natif (**Figure 45b**). La délétion des autres queues d'histones ne semble pas impacter sur l'affinité de l'IN pour le nucléosome, suggérant que la queue d'histone H4 puisse être requise pour un ancrage optimal de l'IN sur le nucléosome. Afin d'évaluer l'impact des queues d'histones sur des complexes d'intégration fonctionnels, des tests d'intégration concertés ont été réalisés sur les différentes constructions de mononucléosomes utilisées précédemment.

Pour quantifier l'impact des queues d'histones sur l'intégration *in vitro*, nous avons réalisé un test d'intégration avec l'IN du VIH-1 et les différentes constructions de MN biotinylés qui sont délétés pour une queue d'histone. A l'issue de la réaction les produits ont été séparés grâce à l'interaction des MN biotinylés avec les billes couplées à la streptavidine et magnétisation (**Figure 46a**). Ainsi, la quantification de la radioactivité de la fraction retenue sur les billes mesure la quantité des produits d'intégration qui ont eu lieu sur du MN. Cette quantification corrèle avec les résultats d'interaction obtenus précédemment. Ainsi, seul le mononucléosome pour lequel il manque la queue d'histone H4 est un moins bon substrat d'intégration (**Figure 46b**) car l'intégration est réduite d'environ 30% sur ce substrat, alors que l'absence des autres queues d'histones n'a pas d'impact significatif.

Ces données confirment que la queue d'histone H4 est importante pour l'interaction optimale de l'IN avec le nucléosome ainsi que pour l'intégration *in vitro*. Pour décortiquer plus en détails l'effet des queues d'histones sur l'intégration *in vitro*, nous avons pour la suite de l'étude utilisé des peptides qui sont issus des différentes queues d'histones H2A, H2B, H3 et H4 afin de les ajouter dans des test d'intégration concertée et de voir leur impact sur l'intégration. Parmi ces peptides histones nous utiliseront plusieurs variants de la queue H4 avec divers degrés de méthylation au niveau de la lysine 20 (H4, H4K20me1, H4K20me2 et H4K20me3) car nous avons observé que ces peptides n'avaient pas tous la même affinité pour l'IN (**Figure 32b**).

3 Impact de l'interaction IN-queue d'histone H4 sur l'intégration *in vitro*

Comme on peut le voir en **Figure 47**, l'addition du peptide H4K20me1 (résidus 11 à 23 de la queue d'histone H4), stimule fortement l'intégration du VIH-1 (environ 3 fois plus d'activité à l'optimum de stimulation) sur ADN nu. Les peptides dérivés des autres queues d'histones (H2A, H2B et H3) ont été utilisés comme contrôles négatifs (nous n'avons pas observé d'interaction de ces peptides avec l'IN du VIH) et de spécificité. Dans ces conditions, seul le peptide H4K20me1 est capable de stimuler l'intégration, et nous n'avons pas observé d'effet significatif des autres peptides. Ces résultats montrent que la stimulation induite par le peptide H4 est spécifique de cette queue d'histone et semble dépendante de l'interaction du peptide avec l'IN. Pour confirmer cette hypothèse, et parce que les peptides sont riches en charges positives (ce qui pourrait interférer avec l'interaction de l'ADN chargé négativement)

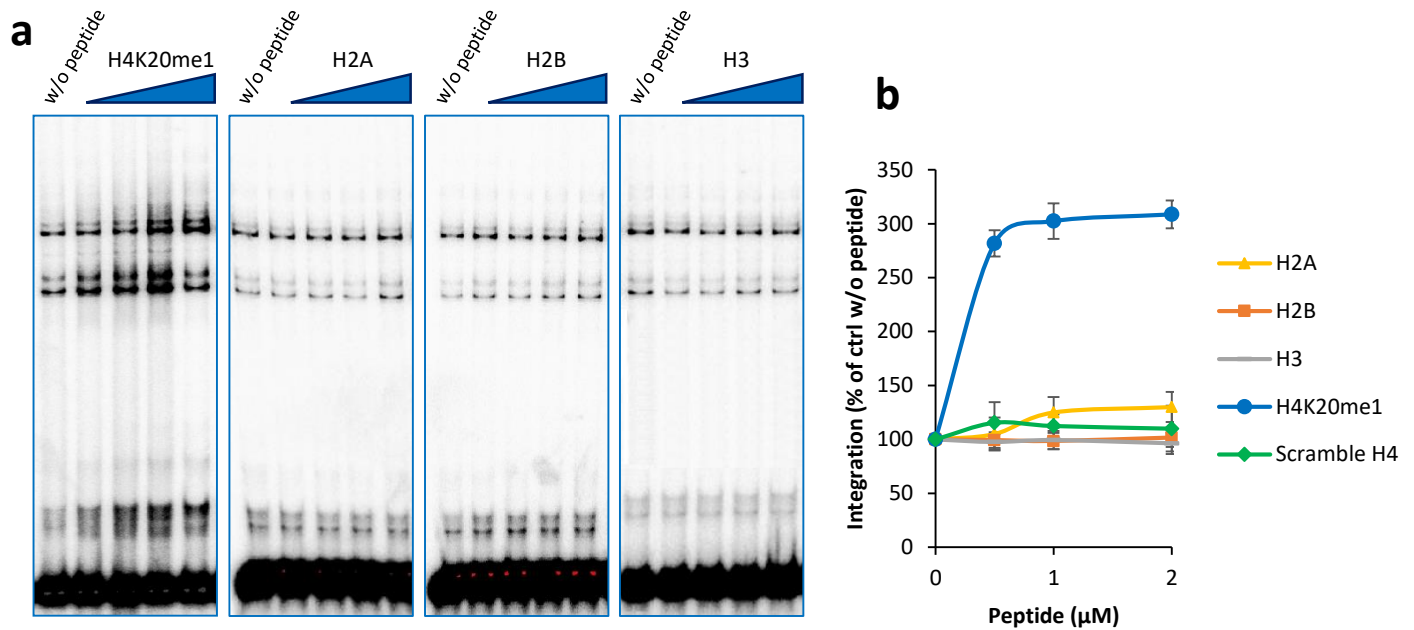


Figure 47 : Influence des peptides dérivés de queues d'histones sur l'intégration *in vitro*. (a) Gels d'intégration concertée réalisés en présence des peptides issus des différentes queues d'histones H4, H2A, H2B et H3 (0, 0,5, 1, 2 et 5μM) et la quantification de l'intégration en (b). (Le scramble H4 partage la même composition en acides aminés que le peptide H4 mais agencés aléatoirement).

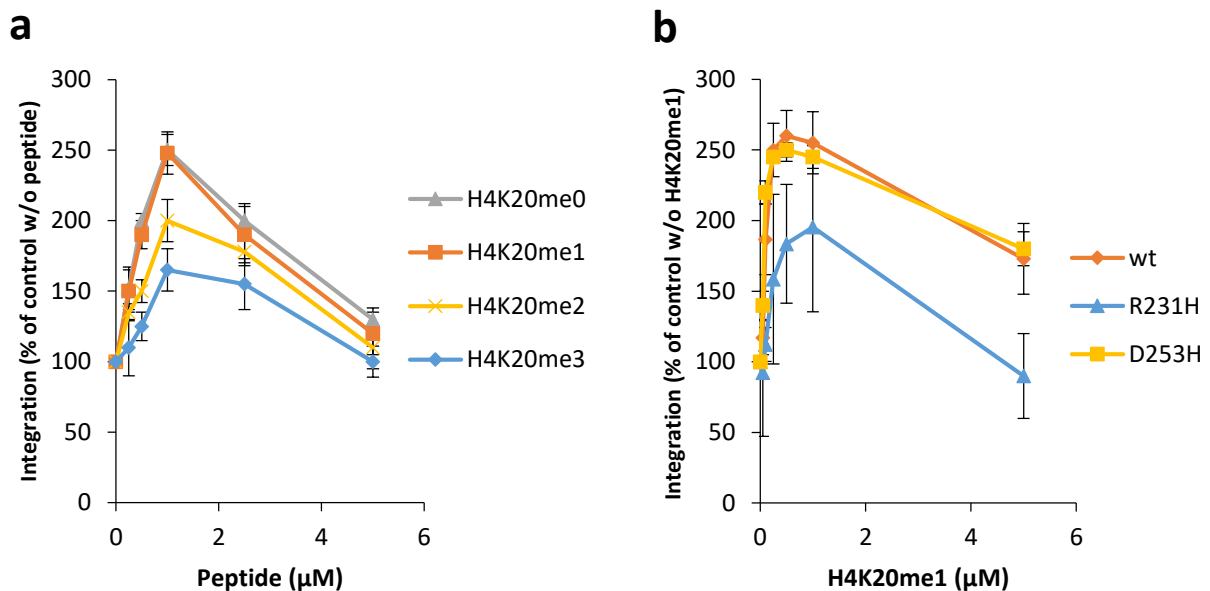


Figure 48 : Influence de l'interaction IN-peptide histone sur l'intégration *in vitro*. (a) Quantification des gels d'intégration concertée réalisés avec des peptides histone H4 ayant divers degrés de méthylation sur la lysine 20. (b) Quantification des gels d'intégration concertée réalisés avec les mutants R231H et D253H en présence du peptide histone H4K20me1. Ces résultats ont été réalisés sur au moins trois expériences indépendantes.

nous avons généré un peptide « scramble » de même taille, de même charge globale et de même composition que le peptide H4 mais dont la séquence primaire a été mélangée aléatoirement. L'addition de ce peptide scramble dans le test d'intégration n'a pas affecté la catalyse (**Figure 47b**) soulignant une fois de plus que la stimulation est spécifique de la queue d'histone H4K20me1 et est probablement reliée à l'interaction IN-peptide.

Pour confirmer cette hypothèse, nous avons mis à profit l'utilisation des différents outils que nous avons développés et caractérisés précédemment, à savoir le fait que l'interaction peut être modulée par l'utilisation des différentes formes méthylées de H4K20 ou même l'utilisation des mutants dont l'interaction avec le peptide H4K20me1 a été décrite dans l'axe précédent. Ainsi, nous avons pu voir que l'addition des peptides avec divers degrés de méthylation n'affectait pas la stimulation d'intégration de la même manière (**Figure 48a**). L'ampleur de la stimulation est en effet dépendante de la capacité du peptide à interagir avec l'IN (e.g. les H4K20me1 et H4K20me0 qui interagissent mieux avec l'IN que les H4K20me2 et H4K20me3 stimulent d'avantage l'intégration comparé à ces derniers). La même tendance a été observée pour le comportement du mutant R231H qui interagit moins avec la queue H4K20me1 par rapport à l'enzyme sauvage, la stimulation est alors plus faible que pour l'enzyme non mutée (**Figure 48b**). Dans le cas du mutant D253H dont l'interaction avec la queue d'histone est meilleure, l'ajout du peptide stimule l'intégration autant que pour l'IN sauvage. Ces observations semblent confirmer notre hypothèse selon laquelle la stimulation de l'intégration semble étroitement liée à la capacité d'interaction entre l'IN du VIH et la queue d'histone H4.

Nous avons par la suite évalué la spécificité de cette stimulation sur différentes intégrases rétrovirales afin de voir si le peptide H4K20me1 pouvait avoir un effet sur d'autres enzymes que celle du VIH. Les intégrases de PFV (prototype foamy virus, spumavirus), de RAV (rouss associated virus, alpharetrovirus) et de MVV (maedi visna virus, lentivirus) ont donc été testées dans le test d'intégration concertée *in vitro* en présence du peptide H4K20me1. Comme on peut le voir en **Figure 49**, ni l'IN de RAV ni celle de PFV ne sont stimulées par l'ajout du peptide H4K20me1. En revanche l'IN de MVV, qui appartient également au genre lentiviral, est aussi stimulée par la queue d'histone H4 bien que plus faiblement que celle du VIH (environ deux fois plus d'activité à l'optimum de stimulation pour MVV contre trois fois plus pour VIH comparé à l'activité basale). Bien que nous n'ayons pas évalué la capacité d'interaction de l'IN de MVV avec le peptide histone H4, il semble cependant qu'il y ait une corrélation entre capacité d'interaction du peptide avec l'IN et capacité à stimuler l'intégration. Bien que cela nécessite de tester et de caractériser d'autres intégrases lentivirales, au vu de ces résultats, il n'est pas exclu que l'implication de la queue d'histone H4 dans l'intégration soit spécifique et/ou conservée au sein du groupe des lentivirus.

En analysant de plus près l'effet du peptide H4K20me1 sur l'intégration du VIH *in vitro*, on remarque que le peptide stimule particulièrement l'hétérointégration par rapport à l'autointégration (**Figure 47a** et quantification des différentes formes en **Figure 50a**). Ceci suggère que le peptide agit sur des complexes de transfert de brin qui vont catalyser préférentiellement de l'hétérointégration. Dans ce test d'intégration *in vitro*, la forme qui se rapproche le plus d'une intégration physiologique telle qu'elle est attendue dans la cellule

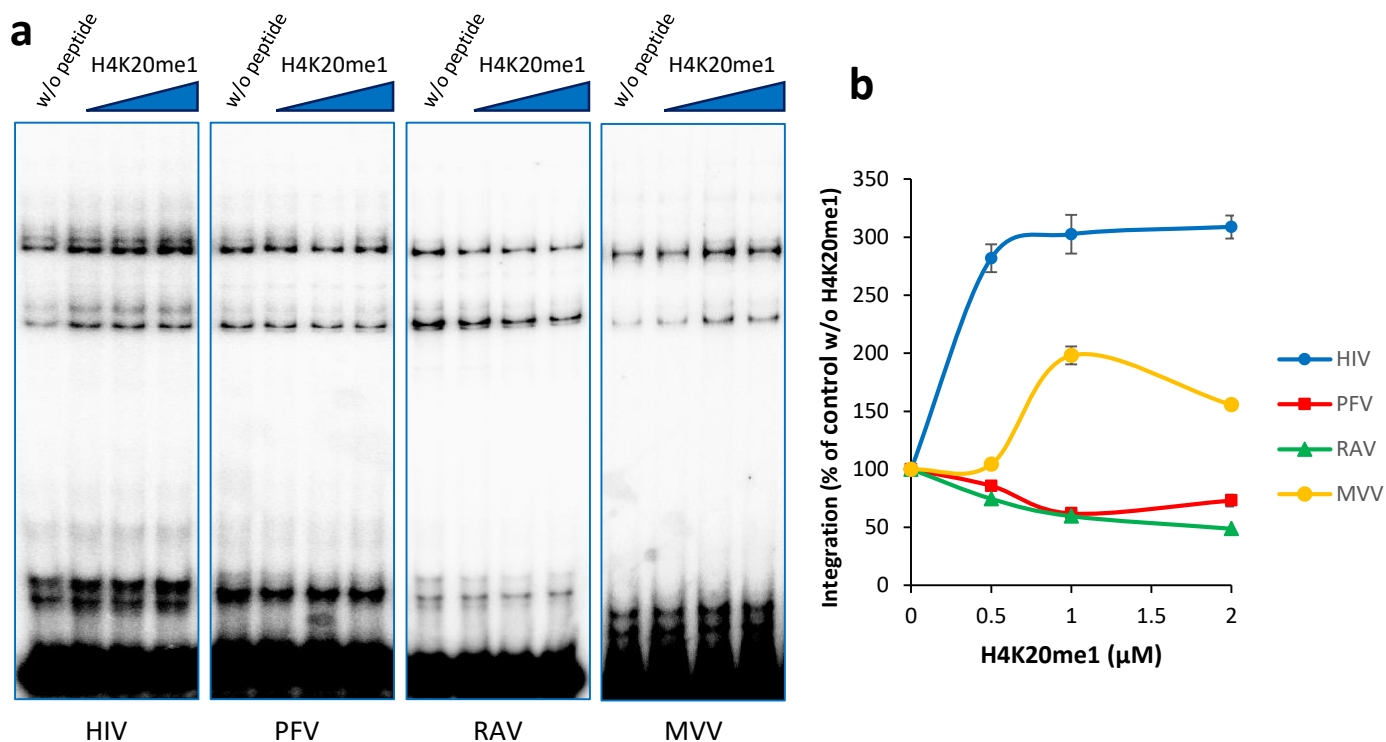


Figure 49 : Influence du peptide histone H4K20me1 sur différentes IN rétrovirales. (a) Gels d'intégration concertée réalisés avec différentes IN (HIV, PFV, RAV et MVV) et en présence du peptide histone H4K20me1 (0, 0,5, 1, 1,5 et 2µM) et leur quantification en **(b)**. Ces résultats ont été réalisés sur au moins trois expériences indépendantes.

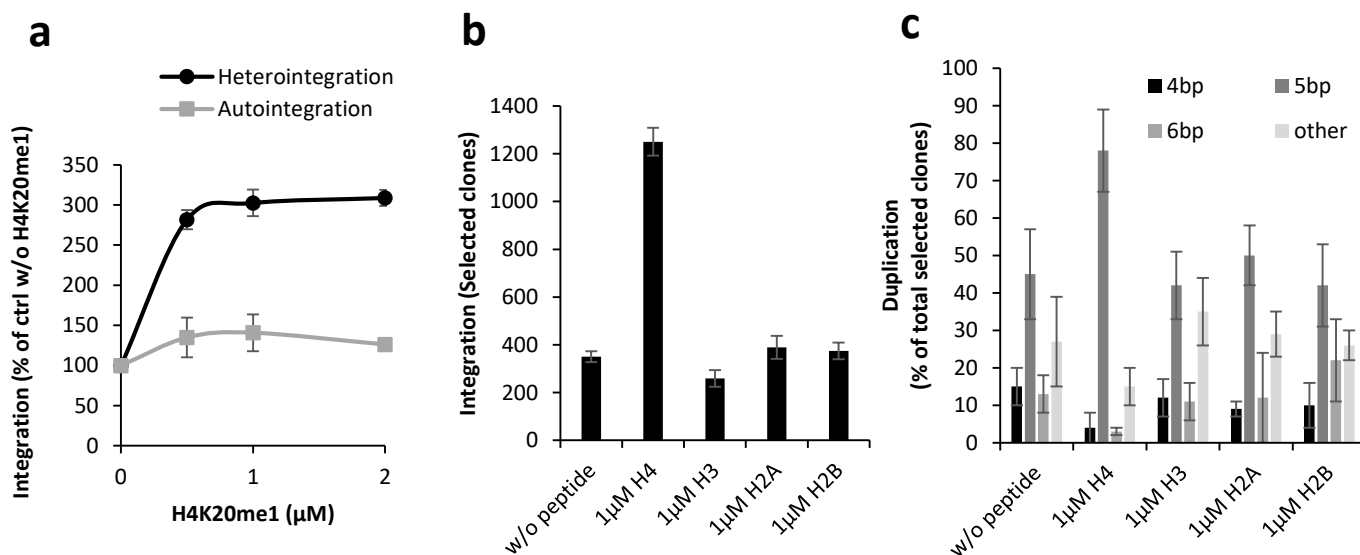


Figure 50 : Influence des peptides histones sur la qualité des produits d'intégration. (a) Quantification des produits d'hétérointégration et d'autointégration induits par le peptide histone H4K20me1 lors d'un test IC. **(b)** Nombre de clones obtenus suite à un test d'intégration non radioactif en présence des différents peptides histones dont les produits ont été transformés dans des bactéries. **(c)** Analyse de la duplication de paires de bases des produits d'intégration clonés dans les bactéries en présence des différents peptides histones. Ces résultats ont été réalisés sur au moins trois expériences indépendantes.

est l'intégration full site circulaire (complète) résultant de l'intégration concertée d'un ADN donneur dans le plasmide receveur. Malheureusement elle n'est pas quantifiable directement avec le test d'intégration *in vitro* car ces produits migrent à la même taille que les produits d'intégration half site circulaires (intégration incomplète d'une seule extrémité virale). Il est cependant possible de cloner les produits d'intégration circulaires issus du test d'intégration *in vitro* (qui est ici réalisé avec un ADN donneur non radiomarké), dans des bactéries et ainsi de quantifier le nombre de clones résistants à divers antibiotiques en présence du peptide ou non. Le résultat des clonages en présence des différentes queues d'histone est présenté en **Figure 50b**, on voit que l'ajout du peptide H4K20me1 augmente le nombre de clones (et donc la proportion d'événements d'intégration full site circulaires) et que cet effet est spécifique du peptide dérivé de cette queue d'histone. Bien qu'on ne sache pas pour le moment si des intasomes catalysant des événements full site présentent des différences structurales par rapport à des intasomes qui eux catalysent de l'intégration half site *in vitro*, il semble néanmoins que le peptide agisse très majoritairement sur les complexes de transfert de brin qui vont catalyser de l'intégration physiologique full site. De plus, le séquençage des sites d'intégration montre un enrichissement en produits contenant une duplication de 5 paires de bases au niveau de l'ADN cible (**Figure 50c**), qui est la duplication retrouvée normalement pour le VIH. L'ensemble de ces résultats semble indiquer que la stimulation induite par le peptide s'exerce préférentiellement sur la capture de l'ADN cible, étant donné que l'hétérointégration est favorisée par rapport à l'autointégration et que la différence majeure entre les complexes de transfert de brin qui catalysent ces produits est la nature de l'ADN cible. Pour continuer de décortiquer l'effet du peptide sur l'intégration, nous avons ensuite testé son effet sur les différentes étapes de l'intégration : le 3' processing et le transfert de brin.

4 Etude *in vitro* du mécanisme d'action du peptide histone H4

4.1 Effet du peptide histone H4 sur le processing et le transfert de brin

Afin de voir si le peptide H4K20me1 a un effet sur une des étapes catalytiques de l'intégration (3' processing et transfert de brin) nous avons dans un premier temps ajouté la queue d'histone dans un test de 3' processing (**Figure 51a**). Dans ces conditions, aucun effet du peptide n'a été observé pour cette étape. Même conclusion pour l'effet de la queue d'histone sur le test de transfert de brin (**Figure 51c**). Ce résultat est surprenant car nous nous attendions à voir un effet sur au moins une des deux étapes, étant donné que les tests de 3'processing et de transfert de brin ont été réalisés dans les mêmes conditions réactionnelles que le test d'intégration concertée. La différence majeure entre ces tests est la longueur et la nature des substrats ADN : pour le test d'intégration concertée, l'ADN donneur mesure 247pb et a le choix de s'intégrer soit dans un plasmide de plusieurs kb soit dans un autre donneur. Pour catalyser la réaction de processing, l'IN n'a pas besoin de plus d'ADN que les 21pb contenues dans le substrat de processing, donc à priori la réaction se fait de la même manière

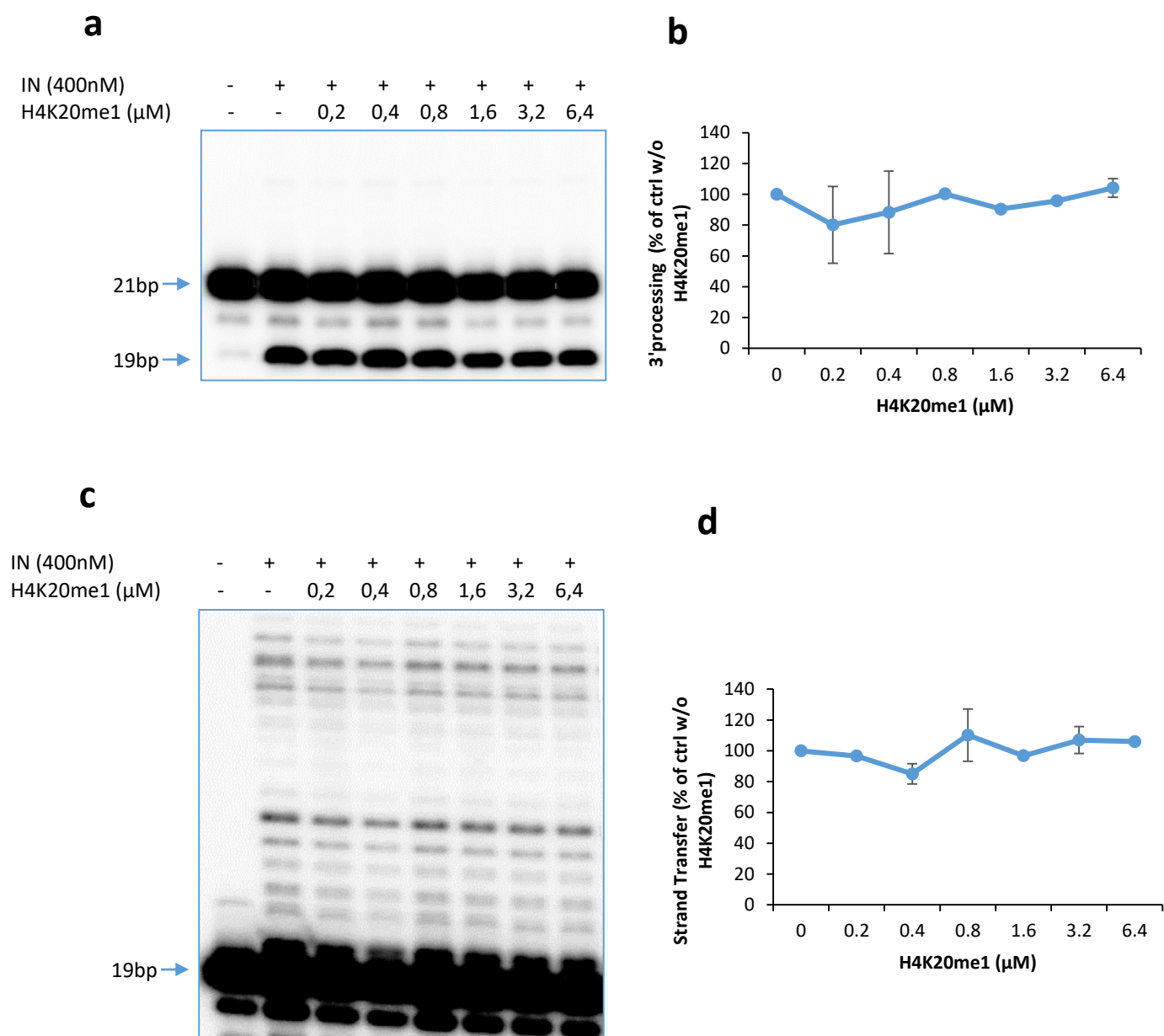


Figure 51 : Effet du peptide histone H4K20me1 sur le processing et le transfert de brin. (a) et (b) Effet du peptide histone H4K20me1 sur la réaction de 3' processing. (c) et (d) Effet du peptide histone H4K20me1 sur la réaction de transfert de brin. Ces résultats ont été réalisés sur au moins trois expériences indépendantes.

que dans le test d'intégration concertée. En revanche pour le test de transfert de brin, l'IN n'a pas d'autres choix que d'intégrer l'ADN donneur dans le fragment d'ADN de 19pb, ce qui mime une réaction d'autointégration plus que d'hétérointégration, or nous avons observé précédemment que le peptide n'avait que peu d'effet sur l'autointégration dans le test d'intégration concertée. Ainsi ces données montrent que le peptide n'a pas d'effet sur le processing ni sur le transfert de brin dans ces conditions.

4.2 Effet du peptide histone H4 sur des intasomes

Il semblerait pour le moment que l'effet stimulateur du peptide H4K20me1 soit dû à un effet sur la capture de l'ADN cible et/ou sur l'étape de transfert de brin. Pour confirmer ces hypothèses, nous avons testé le peptide sur l'intégration *in vitro* en utilisant cette fois-ci des intasomes purifiés. Ceci nous permet de nous affranchir d'un potentiel effet du peptide sur l'assemblage de l'intasome et sur le processing (les ODN utilisés pour l'assemblage sont déjà processés), ainsi seuls la capture de l'ADN cible et le transfert de brin sont observés dans ce test. Pour cela, nous avons assemblé de l'intasome MVV et non du VIH car nous ne savons pas assembler d'intasome VIH actuellement. Il y a tout de même des conditions décrites basées sur une construction de l'IN fusionnée au domaine Sso7d qui augmente drastiquement la solubilité de l'enzyme⁹¹. Etant donné que l'IN est modifiée, il n'est pas exclu que les intasomes résultant de cette construction ne reflètent pas la physiologie de l'intasome VIH qui serait normalement attendue, cependant il serait intéressant de tester l'effet de la queue d'histone H4 sur cette construction. L'intasome de MVV est une bonne alternative à ces contraintes car les conditions d'assemblage de cet intasome lentiviral sont décrites⁹⁰ et que l'intégrase recombinante de MVV est stimulée par le peptide dans le test d'intégration concertée. Nous avons également purifié de l'intasome PFV en tant que contrôle négatif car l'IN recombinante de PFV n'est pas stimulée par la queue d'histone. Les intasomes ont été assemblés et purifiés par gel filtration (**Figure 52a et b**). Les fractions correspondant au pic de l'intasome ont été isolées et utilisées pour un test d'intégration *in vitro* sur plasmide (**Figure 52c**) pour contrôler leur activité. On peut voir que ces intasomes sont très actifs sur plasmide : la bande correspondant à de l'intégration full site (FSI product, qui migre à la taille attendue du plasmide linéarisé) apparaît à des temps très courts. On peut même observer un « smear », qui est la résultante de plusieurs intégrations full site dans le plasmide. Comme le montre la quantification du produit full site (**Figure 52d**), l'ajout du peptide H4K20me1 stimule l'intégration de l'intasome MVV (d'environ 40% à l'optimum de stimulation), ce qui est proche de l'amplitude de stimulation observée avec l'IN recombinante, alors que nous n'avons observé aucun effet sur l'intasome PFV. Ces résultats montrent que le peptide peut stimuler des intasomes déjà formés et suggèrent que cette stimulation est liée à la formation des complexes de transfert de brin médiée par l'interaction IN-H4, c'est la raison pour laquelle nous avons ensuite testé l'effet du peptide H4 sur l'association IN-ADN cible.

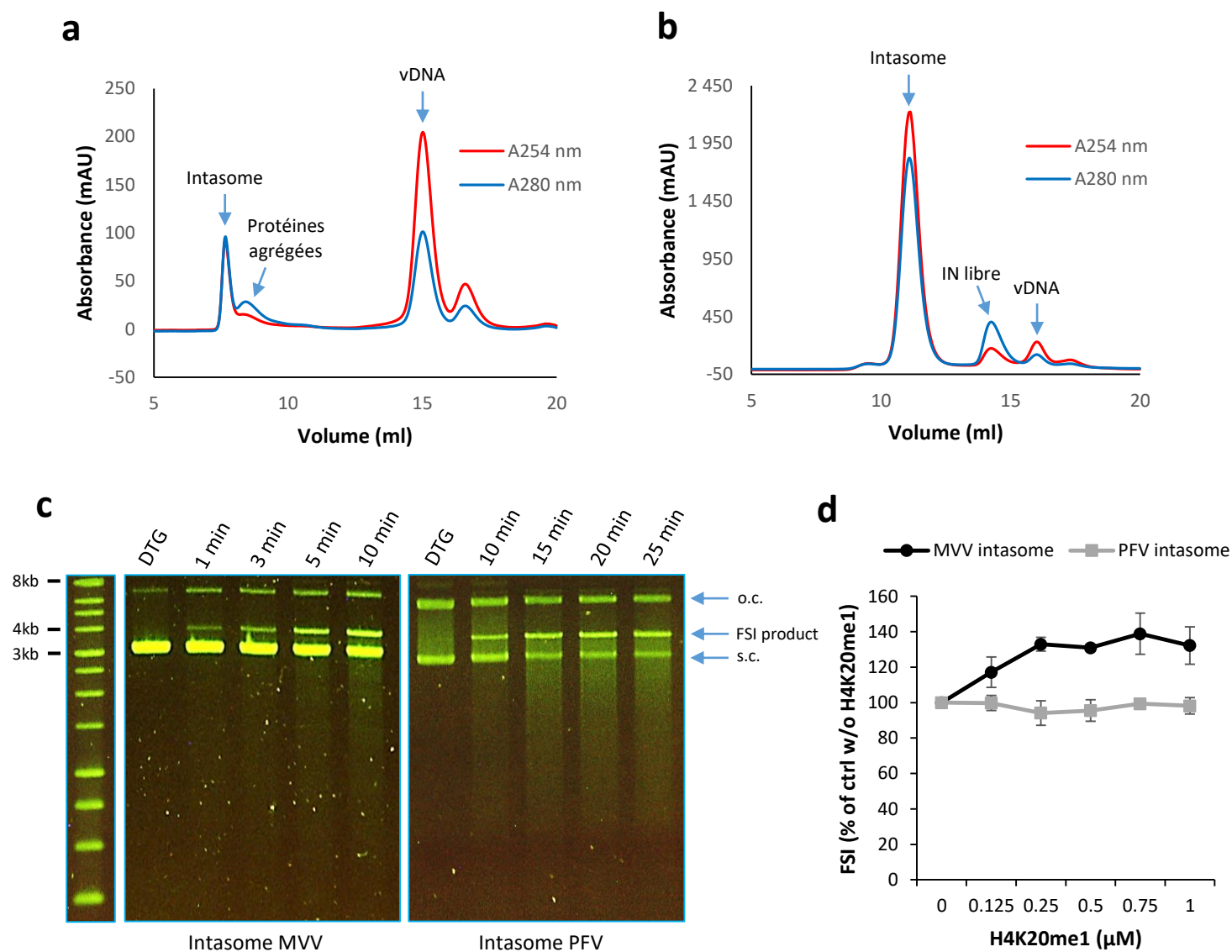


Figure 52 : Influence du peptide histone H4K20me1 sur les intasomes de MVV et PFV. Profil d'élution de l'intasome de MVV **(a)** et de PFV **(b)**. Test d'intégration sur plasmide avec les intasomes purifiés (MVV panel de gauche et PFV panel de droite) en cinétique **(c)**. Quantification de l'effet du peptide histone H4K20me1 sur l'intégration des intasomes de MVV et de PFV **(d)**. Ces résultats ont été réalisés sur deux expériences indépendantes.

4.3 Effet du peptide histone H4 sur la capture de l'ADN

Pour évaluer l'effet du peptide sur l'association entre l'IN et l'ADN cible, nous avons utilisé le fragment d'ADN biotinylé 601 (séquence de 147pb servant à l'assemblage des mononucléosomes, qui mime donc un ADN cible) que nous avons fixé sur des billes magnétiques couplées à la stréptavidine. Nous évaluons ensuite la quantité d'IN recombinante du VIH fixée sur l'ADN en présence ou non du peptide H4K20me1 par western blot anti intégrase. Comme on peut le voir en **Figure 53a**, l'ajout du peptide dérivé de la queue d'histone H4 permet d'augmenter la quantité d'IN du VIH retenue sur l'ADN 601. A nouveau nous avons utilisé l'IN de PFV en tant que contrôle négatif (car l'intégration de PFV n'est pas stimulée par le peptide), et nous n'avons pas observé d'effet du peptide pour cette IN. A ce stade, il semblerait que l'hypothèse selon laquelle le peptide permet de stimuler l'intégration par modulation de l'association IN-ADN cible soit plausible, cependant ces expériences ont été réalisées avec de l'IN recombinante et non des intasomes (il n'existe pas pour le moment d'anticorps contre l'IN MVV), il n'est ainsi pas exclu que ce que nous avons observé ici ne reflète pas ce que nous aurions eu avec des intasomes, car les contacts IN-ADN ne sont pas forcément les mêmes. Cependant, nous avons également réalisé ces expériences de pull down avec un ADN court HU5 utilisé comme ADN donneur dans les tests d'intégration sur MN, et dans ce cas-là, nous n'avons pas observé d'effet significatif du peptide sur la quantité d'IN retenue sur cet ADN (**Figure 53c**). Il semble donc que le peptide augmente l'affinité de l'IN pour l'ADN en favorisant certains contacts uniquement, car les contacts impliqués dans l'interaction avec le LTR ne semblent pas affectés par la queue d'histone (voir pull down avec l'ADN HU5). Cette observation est concordante avec les résultats concernant les intasomes où nous avons vu que le peptide pouvait stimuler un intasome déjà formé, en augmentant/favorisant potentiellement son interaction avec l'ADN cible.

4.4 Effet du peptide histone H4 sur l'oligomérisation de l'IN

Il n'est pas exclu que le peptide puisse également avoir un effet additionnel sur l'assemblage même de l'intasome, et donc sur l'oligomérisation de l'IN. Par ailleurs, l'oligomérisation de l'IN et l'interaction à l'ADN sont étroitement liés, comme nous venons de voir que le peptide histone augmentait l'interaction de l'IN à l'ADN, il est possible qu'il puisse également avoir un effet sur l'oligomérisation. Pour vérifier cela nous avons réalisé des pontages chimiques sur l'IN au BS3 en présence ou non des peptides dérivés des queues d'histones. Comme présenté en **Figure 54a**, on peut voir que le peptide dérivé de la queue d'histone H4 induit une oligomérisation de l'IN visible par l'augmentation des formes correspondantes aux dimères, tétramères et formes de plus haut poids moléculaires. Cet effet oligomérisant du peptide H4 n'a pas été observé pour l'IN de PFV (utilisée en contrôle négatif car elle n'interagit pas avec le peptide H4) (**Figure 54b**). Pour s'assurer de la spécificité de l'effet observé, les autres peptides issus des autres queues d'histones H2A, H2B et H3 ont également été testés, mais seul H4 est capable d'induire une oligomérisation de l'IN du VIH

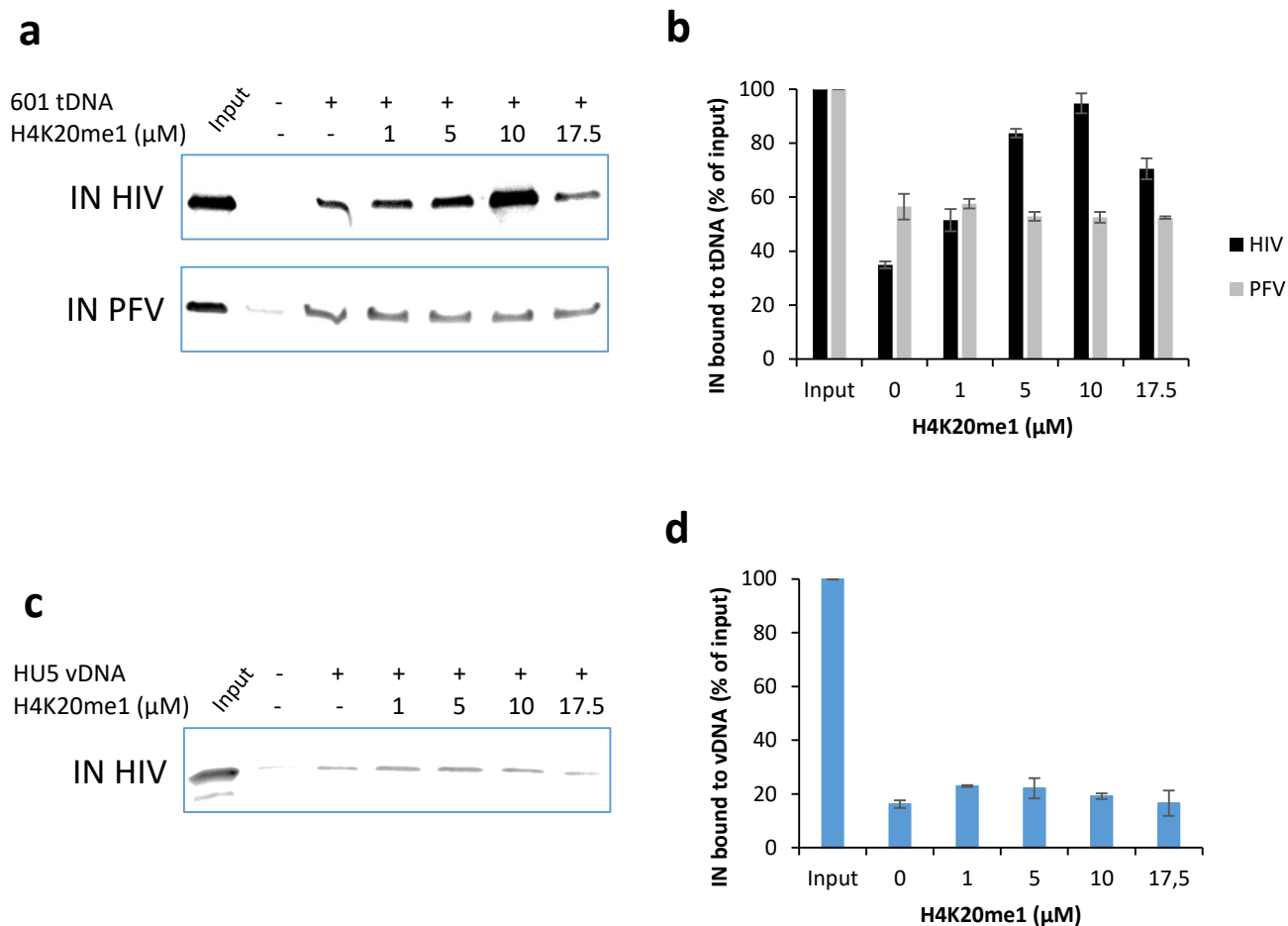


Figure 53 : Effet du peptide histone H4K20me1 sur l'interaction avec l'ADN. (a) Interaction entre l'IN du VIH (panel du haut) ou de PFV (panel du bas) avec l'ADN 601 en présence du peptide histone H4K20me1 par pull down avec leur quantification en **(b)**. **(c)** Interaction entre l'IN du VIH et l'ADN HU5 en présence du peptide histone H4K20me1 par pull down et sa quantification en **(d)**. Ces résultats ont été réalisés sur au moins deux expériences indépendantes.

dans ces conditions (**Figure 54d**). Il semble que cet effet soit à nouveau dépendant de l'interaction entre l'IN et le peptide. A ce stade nous ne savons pas si le peptide H4 induit vraiment une oligomérisation compatible avec le degrés d'oligomérisation des intasomes ou s'il induit une agrégation anarchique de l'IN. Nous pensons que la deuxième hypothèse peut être écartée car si le peptide avait un tel effet délétère sur l'IN, nous n'aurions probablement pas observé une stimulation de l'intégration concertée d'une telle ampleur (**Figure 47**). Pour essayer de valider l'autre hypothèse et montrer que ces oligomères sont actifs pour l'intégration *in vitro* (et que peut être une espèce oligomérique est plus active qu'une autre), nous avons voulu adopter une stratégie similaire à celle utilisée par Faure et al., 2005¹⁸³ où il a été montré qu'il fallait au minimum un tétramère d'IN pour catalyser de l'intégration concertée *in vitro*. La stratégie consiste à ponter l'IN en présence du peptide H4, de purifier indépendamment les différents oligomères et de les tester en intégration concertée *in vitro*. Malheureusement dans ces conditions de pontage, nous n'avons pas réussi à obtenir de l'activité quand l'IN était pontée et la purification des différentes espèces a ainsi été abandonnée.

5 Etude Structurale de l'effet du peptide H4 sur l'IN

L'ensemble de nos données (stimulation de l'intégration, interaction avec l'ADN, oligomérisation...) suggère que l'association du peptide H4 avec l'IN puisse entraîner des réarrangements structuraux de l'IN par son interaction avec le CTD. Pour identifier un potentiel effet du peptide sur la structuration de l'IN, nous avons entrepris une étude structurale par RMN de l'interaction entre le CTD de l'IN (I220-D270) du VIH et le peptide histone H4, en collaboration avec les Dr M.Ruff (IGBMC, Strasbourg) et S.Bouaziz (Laboratoire de Cristallographie et RMN Biologiques, Paris). Le spectre SOFAST-HMQC réalisé avec le CTD seul (sans peptide) montre que la protéine est parfaitement structurée avec des pics de résonance de ¹H et ¹⁵N entre 6,7 et 10 ppm et entre 104 et 130 ppm respectivement, validant les conditions utilisées (**Figure 55a**).

En présence du peptide H4K20me1, le CTD de l'IN est déstructuré, visible par le regroupement de résonances HN autour de 8 ppm (visible dans le cadre rouge, **Figure 55b**), caractéristique des protéines dépliées. Le peptide H4K20me1 semble déstabiliser ce domaine provoquant une déstructuration du CTD, mais d'après le spectre, il semble également y avoir un autre effet, visible par la largeur des pics en haut du cadre rouge qui peuvent correspondre soit à un effet d'oligomérisation du CTD ou soit à des échanges intermédiaires entre différentes conformations du CTD ou des oligomères différents. Cependant il semble également qu'une petite fraction de la protéine reste structurée dans la même conformation qu'en absence de peptide (cadre vert), visible par les pics parfaitement superposables entre les spectres réalisés avec le CTD seul ou CTD + peptide (**Figure 55c**).

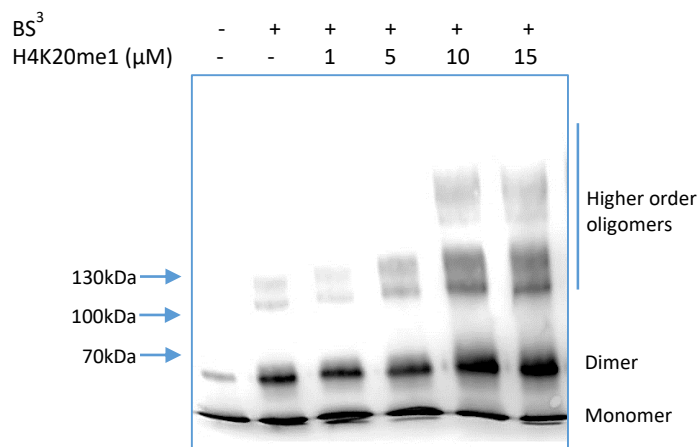
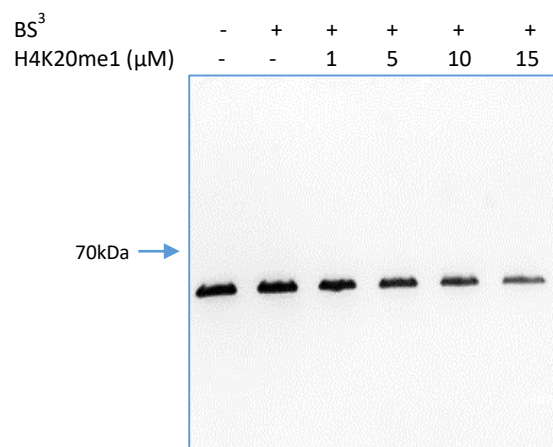
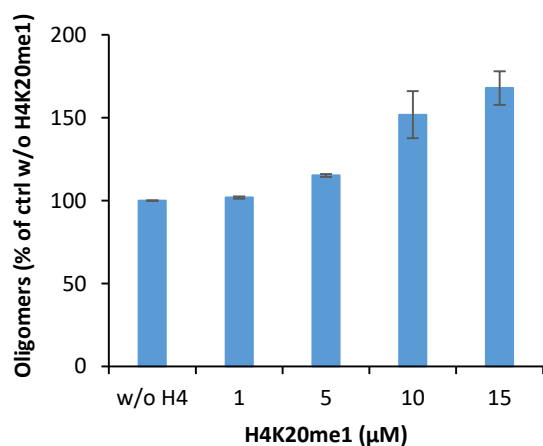
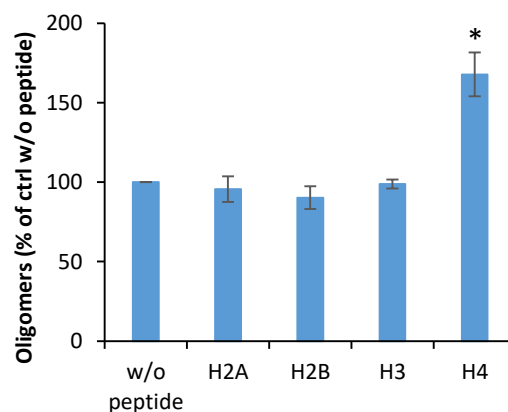
a**b****c****d**

Figure 54 : Effet des peptides histones sur l'oligomérisation de l'IN. (a) Pontage chimique de l'IN du VIH ou de PFV (b) en présence du peptide histone H4K20me1 et la quantification de l'oligomérisation pour VIH en (c). La nature des oligomères du VIH est indiquée à droite du gel. (d) Quantification des pontages chimiques de l'IN du VIH en présence des différents peptides histones (à la concentration de 15μM). Les écarts types ont été calculés avec un test t de Student et représentés par * p<0,05 sur au moins trois expériences indépendantes.

Ces données confirment que le peptide H4K20me1 interagit avec le CTD de l'IN et qu'il est capable d'induire sa déstructuration aboutissant à un effet oligomérisant qui corrèle assez bien avec les données de pontages chimiques réalisés sur l'IN entière. Des expériences de RMN supplémentaires entre le CTD de l'IN et le peptide H4 sont nécessaires, en particulier en utilisant des concentrations croissantes de peptide de manière à effectuer une titration et observer une modification progressive du spectre pour identifier les acides aminés impliqués au cours de cette titration.

6 Conclusion et discussion

L'objectif de cette partie de l'étude était de mieux comprendre le mécanisme moléculaire mis en jeu lors de l'interaction de la queue d'histone H4 avec l'intasome du VIH pendant le processus d'intégration. La résolution de la structure Cryo-EM du complexe intasome de PFV-nucléosome a permis la mise en évidence de contacts directs entre l'intasome de PFV et la queue d'histone H2A. Cette étude a montré que cette interaction est importante pour l'intégration de PFV *in vitro* et *in cellulo* car elle est requise pour l'ancrage optimal de l'intasome sur son nucléosome substrat. Basé sur ces résultats, nous avons également investigué si la queue d'histone H4 pouvait avoir un rôle similaire pour l'IN du VIH.

Par diverses approches *in vitro*, nous avons pu observer que la queue d'histone H4 présente sur les nucléosomes était importante pour l'intégration et l'ancrage optimal du VIH sur des mononucléosomes. Par la suite, l'utilisation des peptides isolés dérivés des queues d'histones nous a permis de mettre en évidence que celui dérivé de la queue d'histone H4K20me1 pouvait stimuler l'intégration du VIH et de MVV. Cette stimulation semble dépendante de l'interaction entre le peptide H4K20me1 et l'IN, car à chaque fois que cette interaction était modifiée (utilisation des variants de méthylation de H4K20me1 ou des mutants d'IN) l'ampleur de la stimulation était également modifiée. De manière intéressante, l'effet stimulateur semble spécifique de la queue d'histone H4 et de l'intégrase VIH (ainsi que MVV). En effet, les contrôles de peptides autres que H4 ou d'IN de différents rétrovirus n'interagissant pas entre eux ne sont pas stimulés. Cependant, nous utilisons un peptide d'une dizaine d'acides aminés pour mimer une queue d'histone qui en fait plusieurs dizaines, ces peptides sont très chargés positivement et il n'est pas exclu qu'ils interagissent non spécifiquement avec à la fois l'IN et/ou l'ADN là où dans la cellule l'interaction se ferait spécifiquement dans une région de la protéine. Un autre facteur qui peut être limitant, est que nous utilisons de l'IN recombinante dans les tests d'intégration concertée et de pull down alors que dans la cellule, c'est de l'intasome qui arrive au contact du nucléosome. Il n'est pas exclu que le comportement d'un intasome par rapport à la queue d'histone H4 soit différent de l'IN recombinante, malheureusement nous n'avons pas pu examiner cela en détails à cause des contraintes techniques que pose l'assemblage de l'intasome du VIH. Nous avons tout de même pu voir qu'avec un intasome MVV, le peptide était capable de stimuler ce complexe.

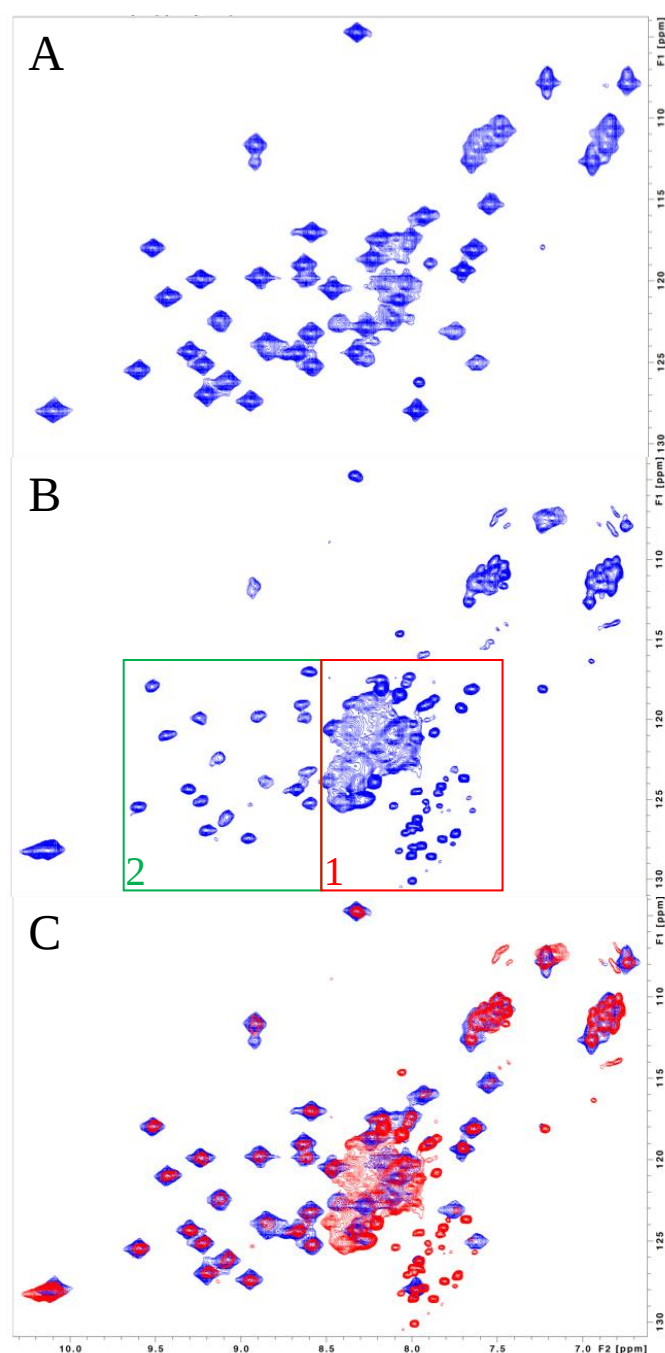


Figure 55 : Effet du peptide histone H4K20me1 sur la structure du CTD de l'IN du VIH. Spectres RMN du CTD de l'IN (résidus 220-270) seul **(a)** ou du CTD et du peptide histone H4K20me1 **(b)**. Le cadre rouge indique la région du CTD qui est déstructurée par le peptide et le cadre vert la région qui reste structurée de la même façon qu'en absence du peptide. **(c)** Superposition des deux spectres RMN (a) en bleu et (b) en rouge.

L'importance de cette stimulation durant le processus d'intégration dans les cellules infectées reste à être pleinement investiguée. Cependant, la baisse de réplication et de l'intégration observée en cellules avec les mutants de l'IN altérés pour leur interaction avec la queue d'histone H4, supporte le fait que cette queue d'histone soit impliquée dans des conditions physiologiques d'infection. L'implication de cette interaction IN- queue d'histone sur la réplication indique qu'il pourrait y avoir un intérêt à développer une stratégie antivirale en ciblant cette interaction, afin d'inhiber l'intégration du virus. De plus, l'effet non négligeable du peptide à promouvoir la formation de complexes d'intégration très actifs fait de ce peptide un bon outil pour des études structure/fonction visant à déterminer le mécanisme par lequel ces complexes se forment ainsi qu'à la résolution de ces structures. En effet, les données de RMN du CTD de l'IN réalisées en présence du peptide H4K20me1 montrent que le peptide est capable d'induire des changements structuraux de ce domaine et que par conséquent, le CTD peut adopter différentes conformations qui dépendent de son interaction avec un partenaire.

Basé sur ces données nous proposons un modèle faisant intervenir la queue d'histone H4 lors du processus d'intégration du VIH dans la cellule (**Figure 56**) : une fois l'intasome arrivé sur son site d'intégration, il va dans un premier temps s'ancrer au nucléosome grâce à des interactions avec l'ADN nucléosomal et la queue d'histone H4. Ensuite, des changements structuraux vont opérer dans l'intasome lui permettant d'interagir optimalement avec l'ADN nucléosomal aboutissant à un état « activé » du complexe de transfert de brin qui va pouvoir catalyser optimalement de l'intégration full site productive pour la réplication du provirus. Bien que ce modèle reste à être confirmé, l'importance de la queue d'histone H4 pour l'intégration du VIH semble une hypothèse solide au vu des résultats accumulés jusqu'à présent, et c'est la raison pour laquelle nous avons choisi de tenter une approche pharmacologique ciblant cette queue d'histone.

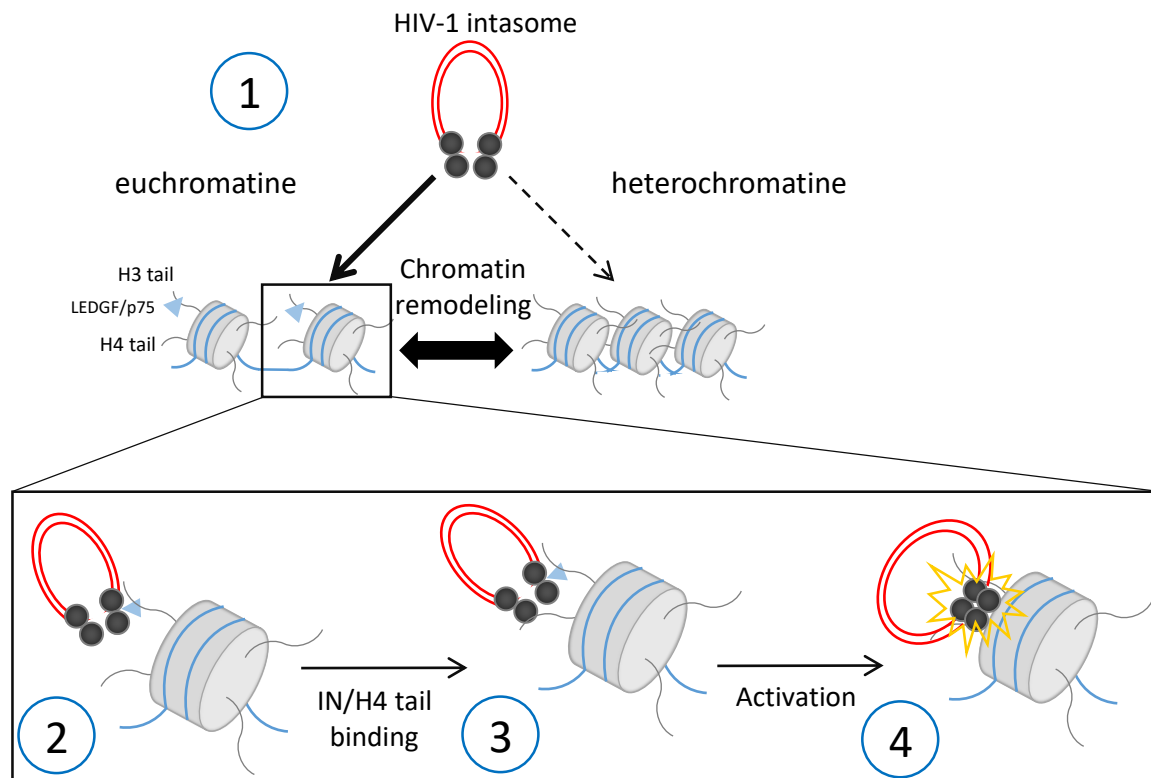


Figure 56 : Modèle d'intégration du VIH-1 faisant intervenir la queue d'histone H4. Suite à son import nucléaire, l'intasome du VIH-1 est guidé dans zones peu denses de la chromatine grâce à CPSF6, et enrichies en LEDGF/p75 **(1)**. La queue d'histone H4 semble avoir un faible impact sur ce ciblage. L'intasome interagit ensuite avec un nucléosome via LEDGF/p75 **(2)** et requiert une association efficace avec la queue d'histone H4 **(3)** pour s'ancrer correctement sur le nucléosome afin de catalyser optimalement l'intégration. De plus nous proposons que cette interaction puisse entraîner des réarrangements de l'intasome **(4)** lui procurant une meilleure affinité pour l'ADN nucléosomal et le rendant plus actif pour catalyser l'intégration.

Partie 3 : L'interaction IN-H4 comme cible antivirale

1 Contexte scientifique et objectifs

L'ensemble de nos données montre que l'interaction entre l'IN du VIH et la queue d'histone H4 semble requise pour que l'intégration soit optimale en cellules. La caractérisation biochimique de cette interaction nous laisse penser qu'elle participe au bon ancrage de l'intasome sur son substrat d'intégration nucléosomal. De plus, nous proposons également que le contact entre l'intasome et la queue d'histone entraîne une « activation » du complexe de transfert de brin par des mécanismes restant à identifier dans le contexte cellulaire semblant impliquer l'affinité pour l'ADN cible ainsi que des réarrangements structuraux au sein du CTD au contact de cette queue d'histone.

Actuellement, bien que les inhibiteurs d'intégration (INSTIs) anti VIH soient extrêmement efficaces et peu toxiques, l'apparition de virus résistants pour un inhibiteur entraîne assez souvent une résistance croisée avec les autres inhibiteurs. Ceci est dû au fait qu'ils se fixent tous dans la même région de l'IN dans le site catalytique, et par conséquent, une mutation dans cette région peut empêcher l'interaction avec tous les INSTIs. C'est une des raisons pour laquelle le développement de nouveaux composés inhibiteurs d'intégration ayant un mécanisme d'action différent est d'intérêt, afin de pouvoir les utiliser même quand le virus aura développé une résistance aux INSTIs actuels.

L'objectif de cet axe était donc d'identifier des molécules chimiques capables d'empêcher l'interaction entre l'IN du VIH et la queue d'histone H4, afin de voir si de tels composés peuvent également inhiber l'intégration du virus en cellules. Ceci nous permettrait non seulement de valider l'interaction IN-H4 comme une nouvelle cible antivirale, mais ouvrirait aussi des portes pour le développement de nouveaux inhibiteurs d'intégration ayant un tout nouveau mécanisme d'action. De plus, ces molécules pourraient également servir d'outils pour mieux appréhender l'effet de cette interaction à la fois *in vitro* et *in vivo*.

La stratégie employée a été dans un premier temps de chercher des molécules pouvant inhiber l'interaction IN-queue d'histone H4 *in vitro*, en criblant une banque de molécules. La seconde étape a été de valider ces molécules *in vitro*, pour voir si elles étaient capables d'inhiber l'interaction et l'intégration sur des mononucléosomes. Ces molécules ont ensuite été caractérisées en cellules afin d'évaluer leur toxicité et leur efficacité sur la réplication virale.

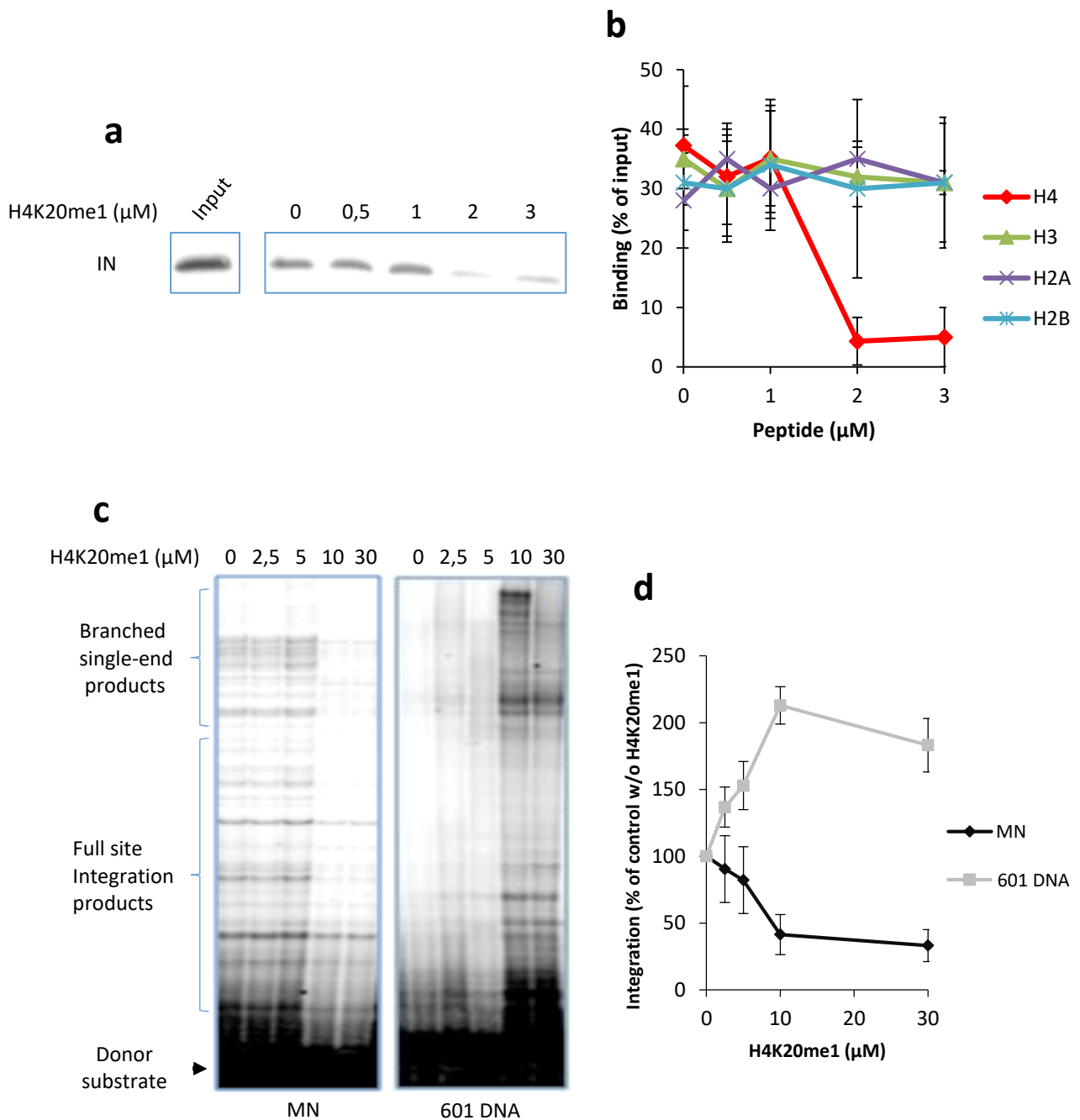


Figure 57 : Effet des peptides histones sur l'interaction et l'intégration *in vitro*. (a) Interaction entre l'IN (1 μM) et le MN biotinylé en présence du peptide histone H4K20me1 par pull down. (b) Quantification de l'interaction entre l'IN et le MN en présence des différents peptides histones. (c) Intégration sur MN (panel de gauche) ou l'ADN 601 (panel de droite) en présence du peptide histone H4K20me1 et la quantification en (d). Ces résultats ont été réalisés sur au moins deux expériences indépendantes.

2 Etude du peptide H4K20me1 comme potentiel compétiteur de l'interaction entre l'IN et la queue d'histone H4

Avant de nous lancer dans le criblage d'une banque de molécules, le peptide dérivé de la queue d'histone H4 a été étudié. En effet, nous nous sommes demandés si ce peptide pouvait entrer en compétition avec la queue d'histone H4 des nucléosomes, prévenant ainsi l'association de l'IN avec ce dernier. Pour cela, nous avons suivi l'interaction entre l'IN et des mononucléosomes biotinylés par pull down, en présence ou non du peptide H4K20me1. Comme on peut le voir en **Figure 57a**, l'ajout du peptide H4K20me1 entraîne une diminution de la quantité d'IN retenue sur les mononucléosomes. Cet effet n'a été observé qu'en présence du peptide dérivé de la queue H4, car les autres peptides dérivés des queues H2A, H2B et H3 n'ont pas d'effet significatif sur l'interaction IN-monucléosome (**Figure 57b**). Ces résultats confirment à nouveau l'importance et la spécificité de l'interaction IN-H4 dans l'association de l'IN au nucléosome, et valident aussi le fait que cette interaction puisse être modulée pharmacologiquement.

Etant donné que jusqu'à présent tous les cas de figures où l'interaction entre l'IN et le nucléosome conduisaient à une inhibition de l'intégration à la fois *in vitro* et *in vivo*, la suite logique a ainsi été de voir si le peptide H4K20me1 pouvait inhiber l'intégration sur mononucléosome. Comme on peut le voir en **Figure 57c**, l'addition du peptide inhibe l'intégration sur mononucléosome, et pour s'assurer que ce n'est pas un effet délétère du peptide sur l'IN dans ces conditions mais bien une inhibition liée au nucléosome, nous l'avons également testé sur la séquence d'ADN 601 qui sert à l'assemblage des nucléosomes (**Figure 57c**). Non seulement le peptide n'inhibe pas l'intégration sur l'ADN 601, ce qui signifie que l'inhibition observée précédemment est bien dépendante du nucléosome, mais il la stimule comme attendu, compte tenu de l'effet du peptide sur les complexes de transfert de brin. En résumé, l'effet du peptide dérivé de la queue d'histone H4 est double sur l'intégration *in vitro* : en l'absence de nucléosomes il est capable de stimuler l'intégration (probablement en favorisant la formation de STC comme nous l'avons vu dans l'axe précédent), alors qu'en présence de nucléosomes l'effet inhibiteur est dominant et empêche la stimulation (probablement en empêchant l'interaction de l'intasome avec le nucléosome et donc la formation du STC). Vu que dans les cellules l'intégration se fait dans un contexte chromatinien, il est possible que le peptide puisse avoir la capacité d'inhiber l'intégration *in vivo*, c'est ce que nous avons cherché à savoir dans la suite du projet.

3 Effet des peptides histones sur les phases précoces de la réplication virale

Afin de déterminer l'effet des peptides issus des queues d'histones sur les phases précoces de la réplication virale, nous les avons transfectés dans des cellules 293T au préalable

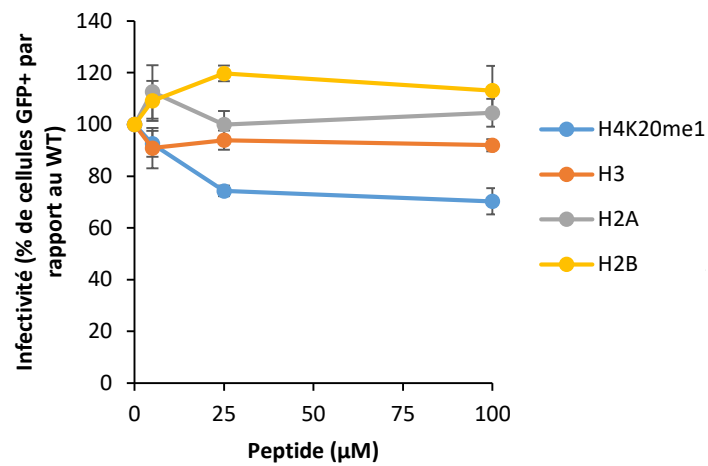


Figure 58 : Influence des peptides histones sur les phases précoces de la réplication. Des cellules 293T ont été transfectées par les différents peptides histones puis transduites par les vecteurs lentiviraux. L'infectivité a été mesurée 10 jours après en suivant l'expression du nombre de cellules GFP positives par cytométrie en flux. Ces résultats ont été réalisés sur deux expériences indépendantes.

d'une transduction par les vecteurs lentiviraux pseudotypés VSVG. Les phases précoces de la réplication ont été analysées par mesure de l'expression du gène rapporteur eGFP par cytométrie en flux. Parmi les peptides transfectés, seul celui dérivé de la queue d'histone H4 a été capable d'induire une baisse de l'infectivité (**Figure 58**). Les peptides ne sont pas toxiques pour les cellules (non montré) mais nous ne savons pas s'ils rentrent efficacement dans la cellule et même s'ils parviennent jusqu'au noyau. Ces résultats suggèrent tout de même que le peptide H4K20me1 puisse bloquer les phases précoces de la réplication virale, en revanche, il faut des quantités élevées de peptide pour atteindre au final un taux d'inhibition globalement assez faible. De plus, il n'est pas exclu que le peptide soit modifié par des méthyl transférases cellulaires en H4K20me2 ou me3, ce qui diminuerait probablement l'affinité du peptide pour l'IN et serait par conséquent un moins bon compétiteur pour l'interaction avec des nucléosomes portant des modifications H4K20me1. Nous n'avons pas exploré plus en détails le mécanisme par lequel le peptide H4 est capable d'inhiber les phases précoces de la réplication en raison de son faible effet. Cependant, ces résultats nous ont conforté dans l'idée d'utiliser l'interaction IN-H4 comme cible antivirale par des composés chimiques.

4 Recherche de composés inhibant l'interaction entre l'IN et la queue d'histone H4

Deux stratégies ont été employées pour identifier des molécules empêchant l'interaction entre l'intégrase et la queue d'histone H4. La première a été de rationaliser des molécules par rapport à la zone d'interaction putative située dans le c-terminal de l'IN que nous avons identifié par modélisation moléculaire. Ce criblage « virtuel » a été réalisé en collaboration avec P.Gouet et X.Robert (IBCP, Lyon), et a pour objectif de sélectionner des molécules dans des banques de données qui peuvent s'agencer dans la zone d'interaction avec le peptide H4 afin d'empêcher ce dernier de venir s'y fixer. La deuxième stratégie est basée sur un crible également, mais un crible physique cette fois-ci. Ainsi, en collaboration avec S.Taouji (institut Bergonié, Bordeaux), nous avons utilisé la méthode alphascreen qui permet de mesurer des interactions entre plusieurs partenaires (dans notre cas entre le C-terminal de l'IN et le peptide H4K20me1) et de tester des banques de molécules à grande échelle, afin d'en identifier certaines qui pourraient empêcher l'interaction IN-H4K20me1. Ces deux stratégies ont été menées en parallèle et seront présentées de manière indépendante avec dans un premier temps le crible virtuel puis le crible physique alphascreen.

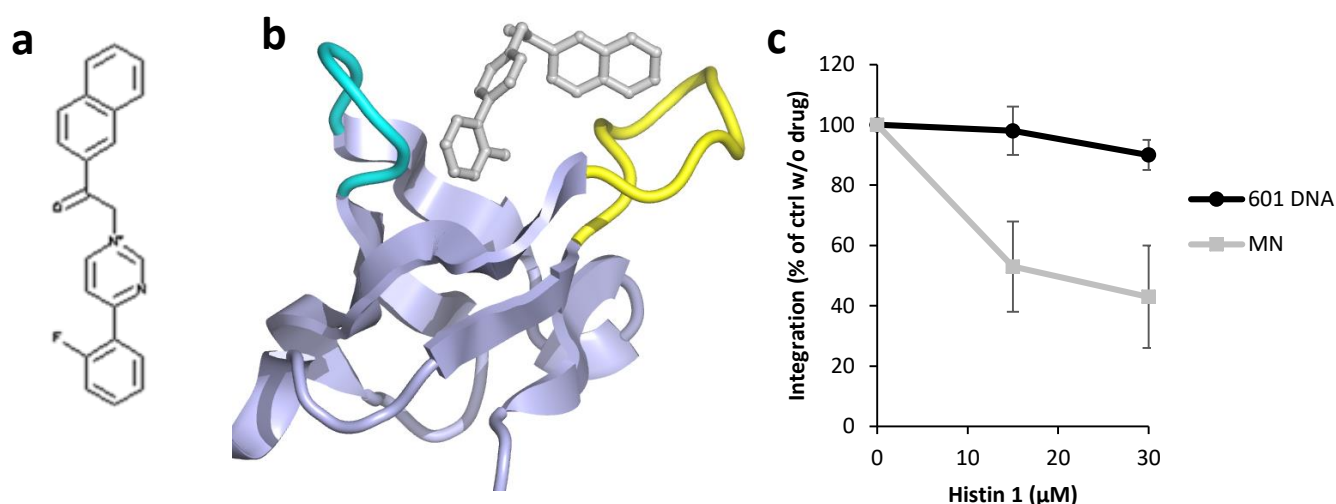


Figure 59 : Analyse du composé ayant eu le meilleur score à l'issu du criblage virtuel avec le CTD de l'IN. (a) Structure chimique du composé. **(b)** Modélisation de l'interaction entre le composé (en gris) et le CTD de l'IN (en violet). La boucle 228-235 est colorée en jaune et la boucle 253-257 en cyan. **(c)** Quantification de l'intégration *in vitro* sur MN et ADN 601 en présence du composé. Ces résultats ont été réalisés sur deux expériences indépendantes.

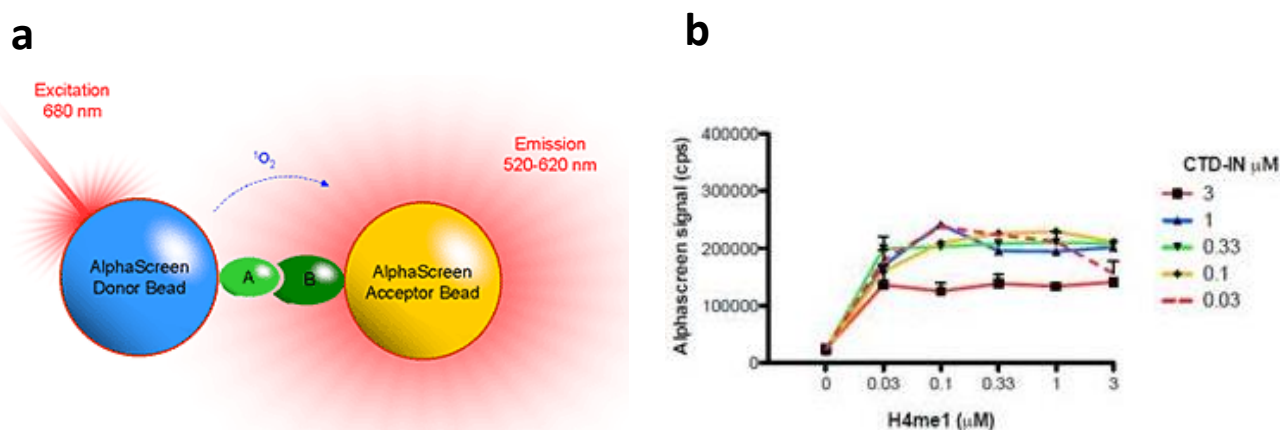


Figure 60 : Mise au point des conditions alphascreen pour suivre l'interaction entre le CTD de l'IN et le peptide histone H4K20me1. (a) Schéma du principe d'action de l'alphascreen. **(b)** Quantification du signal alphascreen suite à l'utilisation de divers ratios du CTD de l'IN et du peptide histone H4K20me1.

4.1 Le criblage virtuel

Le criblage virtuel a été réalisé en collaboration avec X.Robert et P.Gouet, avec comme cible le CTD de l'intégrase du VIH (code PDB : 1EX4). Le sous-ensemble 'Leads-now' de la chimiothèque virtuelle ZINC12 (<http://zinc.docking.org>) a été utilisé comme source de petites molécules candidates. Cette base de données répond partiellement aux critères de Lipinski, ces composés ont donc déjà certains prérequis pour être des molécules à visées pharmacologiques. De plus, la grande majorité sont disponibles commercialement et sont de tailles raisonnables, leur permettant d'être modifiées chimiquement de manière à potentialiser leur effet si besoin. Ces molécules ont été préparées avec le logiciel AutoDockTools. Les 3,687,621 composés de la chimiothèque ont alors été testés grâce au logiciel Autodock/Vina sur la grappe de calcul du laboratoire "Rétrovirus et Biochimie Structurale" (UMR5086, IBCP, Lyon). Dans cette première étape, la protéine cible a été traitée comme rigide et les composés comme flexibles. En conservant les meilleurs scores de docking, ceci nous a conduit à isoler environ 300,000 composés qui ont été soumis à une deuxième phase de docking, cette fois-ci en considérant la cible protéique comme partiellement flexible au niveau de la zone d'interaction considérée. Pour ce faire, le logiciel Autodock/Vina a été à nouveau employé, ce qui a permis d'isoler au final 10 molécules candidates pouvant venir se fixer au niveau de la cavité où le peptide est sensé interagir. Un exemple est montré en **Figure 59b** où on peut voir l'agencement du composé ayant eu le meilleur score (Histin 1) sur le CTD de l'IN parmi les 10 sélectionnés. Certains de ces composés sont en cours d'étude au laboratoire, notamment pour analyser leur capacité à inhiber l'interaction et l'intégration sur mononucléosomes. Bien que les résultats soient encore préliminaires, il semblerait que le composé ayant eu le meilleur score soit capable d'inhiber l'intégration sur mononucléosomes *in vitro* sans affecter l'activité catalytique sur plasmide (**Figure 59c**) bien qu'avec un effet assez modéré car il faut plusieurs dizaines de micromolaires du composé pour inhiber l'intégration de manière efficace. Le reste des molécules de la liste sera également analysé pour leur effet sur l'intégration *in vitro* ainsi que sur les phases précoces de la réplication *in cellulo* en perspective de ce projet.

4.2 Le criblage par alphascreen

L'autre méthode de criblage par alphascreen est basée sur l'utilisation de billes dites « acceptrice » et « donneuse » qui sont fixées à chacun des partenaires comme illustré en **Figure 60a** (dans notre cas le CTD de l'IN est exprimé avec un tag GST qui se fixera sur les billes couplées à la glutathione ; et le peptide biotinylé lui se fixera sur des billes couplées à la streptavidine). S'il y a interaction entre les deux partenaires, cela va permettre aux billes acceptrices et donneuses d'être assez proches pour pouvoir faire une réaction chimique entre elles : les billes donneuses sont excitées à une certaine longueur d'onde et vont émettre une molécule d'O₂ excitée qui pourra réagir avec les billes acceptrices si ces dernières sont assez proches. Les billes acceptrices vont ainsi utiliser cet O₂ afin d'émettre une lumière qui va

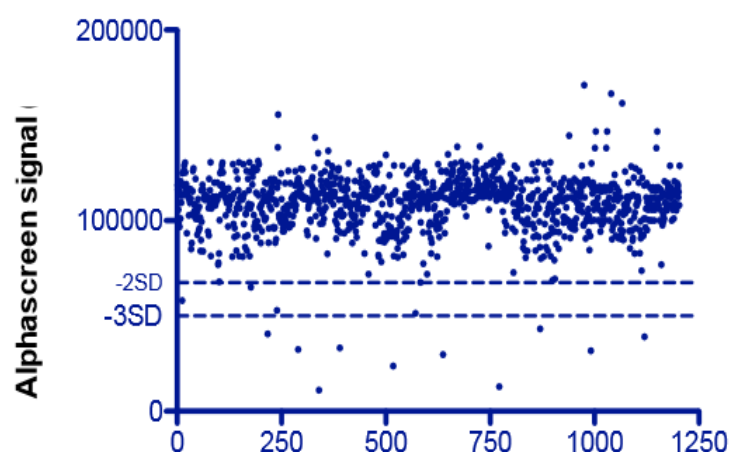


Figure 61 : Recherche de composés inhibant l'interaction entre le CTD de l'IN et le peptide histone H4K20me1. La banque de molécules Prestwick (1280 molécules) a été criblée en utilisant les conditions d'interaction mises au point précédemment. Les composés inhibant significativement le signal en dessous de -3SD (plus de 80% de baisse du signal) ont été sélectionnés pour la suite de l'étude.

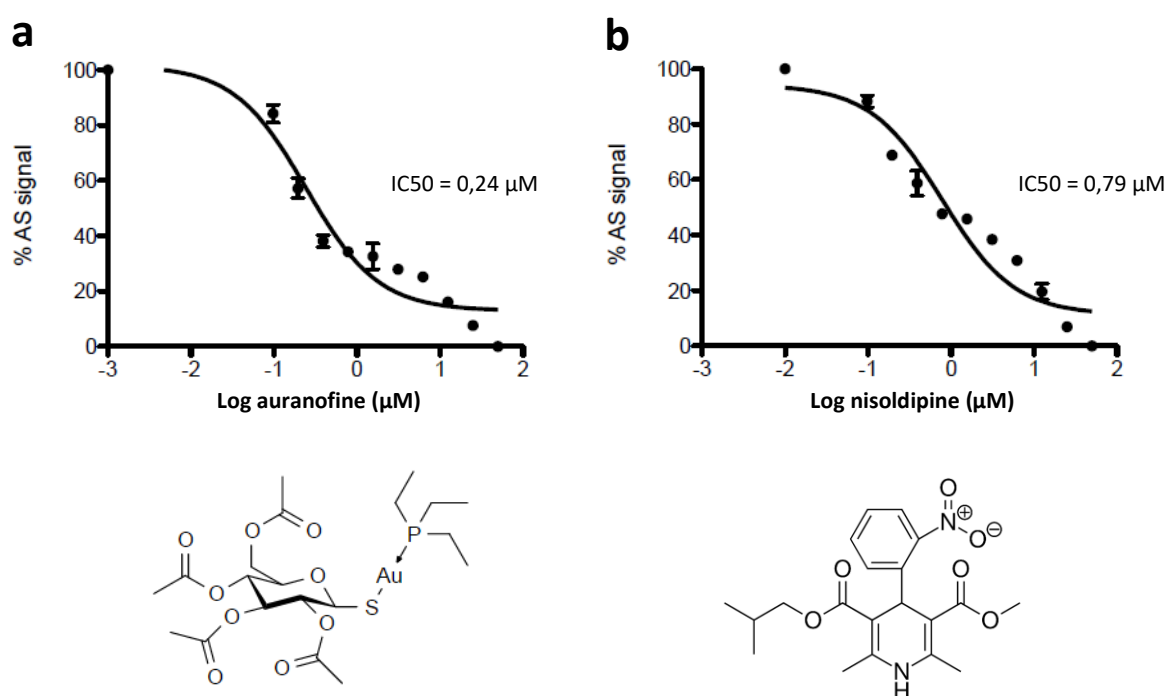


Figure 62 : Analyse de l'efficacité des molécules issues du crible alphascreen à inhiber l'interaction entre le CTD de l'IN et le peptide H4K20me1. L'alphascreen a été utilisé pour calculer les IC_{50} des deux composés en utilisant une gamme de concentration de ces derniers avec en **(a)** l'auranofine et en **(b)** la nisoldipine. La structure chimique des composés est présentée sous les graphiques correspondants.

exciter un fluorophore présent sur ces billes, produisant de la lumière à une longueur d'onde qui sera mesurée et quantifiée.

Avant de commencer le criblage, les conditions ont été mises au point par S.Taouji (plateforme BMyscreen, Bordeaux) pour pouvoir mesurer l'interaction entre le CTD de l'IN et le peptide H4K20me1. Pour cela différents ratios de chaque partenaire ont été testés afin de suivre l'évolution du signal alphascreen, représentatif de l'interaction (**Figure 60b**). Le signal détecté est dépendant de la concentration des partenaires et est proche du bruit de fond quand un des deux n'est pas présent, ce qui valide une interaction spécifique et nous permet donc de cribler une banque de molécules dans ces conditions. Nous avons aussi tenté de mettre au point les conditions d'interaction avec de l'IN entière, malheureusement ce fut infructueux à cause des nombreuses interactions non spécifiques entre l'IN et les billes, le crible a donc été réalisé sur le domaine minimal d'interaction avec la queue d'histone : le domaine C-terminal de l'IN.

Le composé inhibiteur idéal pour une interaction hôte-pathogène doit répondre à plusieurs critères dont une biodisponibilité élevée ainsi qu'une faible toxicité. C'est la raison pour laquelle nous avons choisi la banque de molécules Prestwick qui est composée de 1280 molécules qui sont toutes approuvées par la FDA, et dont les critères cités plus haut sont généralement compatibles pour une administration chez l'Homme. Mr Taouji a réalisé le criblage à un ratio fixe de CTD-IN et de peptide H4 permettant d'avoir un signal confortable pour voir des variations dans ce signal (30nM de CTD et 100nM de peptide H4K20me1). Les composés ont également été testés à une concentration fixe de 10µM. L'analyse du criblage de cette banque de molécules se fait en comparant le signal alphascreen correspondant aux points de référence sans aucun composé, à ceux qui contiennent une des molécules de la banque. On peut voir en **Figure 61** que la majorité des composés n'a pas d'effet significatif sur l'interaction CTD-IN et peptide H4K20me1 (signal autour de 100 000). En revanche certaines molécules diminuent le signal alphascreen significativement, ce qui veut dire qu'elles diminuent l'interaction entre les deux partenaires, et ce sont ces molécules qui nous intéressent en particulier. Ici, nous avons sélectionné les molécules qui baissent le signal de plus de 80% (en dessous de trois fois la différence entre la moyenne et la déviation standard), ce qui correspond pour le moment à 10 molécules. A ce stade, il faut s'assurer que les molécules n'interfèrent pas aspécifiquement avec le signal alphascreen (fluorescence intrinsèque des composés, modification de l'interaction entre les partenaires et les billes), pour cela les contrôles appropriés ont été réalisés et deux molécules ont été sélectionnées comme diminuant spécifiquement l'interaction entre le CTD de l'IN et le peptide H4K20me1. L'IC50 (concentration pour inhiber 50% de l'interaction) a été calculée pour les deux composés (**Figure 62**), (0,2µM pour l'auranofine et 0,8µM pour la nisoldipine).

5 Description des molécules identifiées en alphascreen

L'auranofine (également appelé ridaura) est un complexe d'or organique utilisé pour le traitement de la polyarthrite rhumatoïde. Son mécanisme d'action n'est pas connu, mais certaines fonctions ont été identifiées pour ce composé : il inhibe la libération de certains médiateurs pro inflammatoires, inhibe la kinase cellulaire I κ B, la thioredoxine réductase 1 et a des effets anti-cancéreux. De manière très intéressante, l'auranofine a déjà été un candidat utilisé dans des études thérapeutiques contre le VIH en 2013¹⁸⁴. Cette étude visait à détruire les cellules réservoirs infectées par le VIH de manière latente. Leur stratégie était d'exploiter le fait que ces cellules réservoirs (majoritairement des lymphocytes T mémoires centraux et transitionnels) ont moins de défenses anti-oxydantes comparé aux cellules non différenciées. L'auranofine ayant des effets pro-oxydant en inhibant la thioredoxine réductase, ils se sont demandé si elle pourrait sélectivement détruire les cellules réservoirs du VIH en induisant un burst oxydatif chez ces dernières. Les résultats de cette étude ont été très encourageants, l'auranofine détruisant spécifiquement les cellules mémoires centrales et transitionnelles. De plus, une autre étude menée sur des singes modèles pour l'infection par le VIH a montré que l'auranofine était capable de réduire le réservoir viral, et qu'après arrêt de la thérapie antirétrovirale, seuls les singes ayant été traités avec de l'auranofine présentaient un délai plus important dans la reprise de la réplication virale. Cependant, l'effet de l'auranofine sur la réplication virale en absence de thérapie n'a pas été testé, et leur étude montre que la combinaison thérapie + auranofine diminue encore plus efficacement la détection de l'ADN viral dans le sang des singes (jusqu'à un niveau indétectable) que la thérapie seule. Ce point n'a pas été discuté en détails, car ils le relient à la mort des cellules réservoirs ayant un provirus, mais ces résultats suggèrent également que l'auranofine puisse avoir des propriétés antivirales. Plus récemment, une équipe a proposé que l'auranofine puisse être un inhibiteur de l'entrée du VIH^{185,186}. Lors de l'attachement de la gp120 avec les récepteurs et corécepteurs cellulaires, la gp120 subit des changements conformationnels par la réduction de certains ponts disulfures. Il y a plusieurs enzymes identifiées comme participant à la réduction de la gp120 dont la thioredoxine réductase 1, une enzyme inhibée par l'auranofine. Ainsi dans leur étude, ils ont testé plusieurs composés inhibiteurs de la thioredoxine réductase 1 dont l'auranofine pour voir s'ils étaient capables d'inhiber l'entrée du VIH dans les cellules. Bien que certains de leurs composés bloquent bien l'entrée du virus, validant le fait que la thioredoxine réductase puisse être utilisée comme une cible antivirale, ils ne concluent pas sur l'effet de l'auranofine bien qu'elle ait la capacité à inhiber le virus selon eux (ces données ne sont pas montrées dans leur étude). En résumé, il semblerait que l'auranofine soit déjà à l'étude pour la lutte contre le VIH, en tant que molécule anti réservoir (détruisant spécifiquement les lymphocytes mémoires) et potentiellement en tant qu'inhibiteur d'entrée (en empêchant la réduction de la gp120 et ses changements conformationnels nécessaires pour l'interaction avec les corécepteurs cellulaires). Bien que ces effets ne soient pas directement prouvés, ces études montrent quand même que l'auranofine a la capacité d'inhiber la réplication virale. Notre crible l'a identifiée sur la base de l'interaction entre le CTD de l'IN et le peptide H4K20me1, au vu du peu de preuves sur son mécanisme d'action par

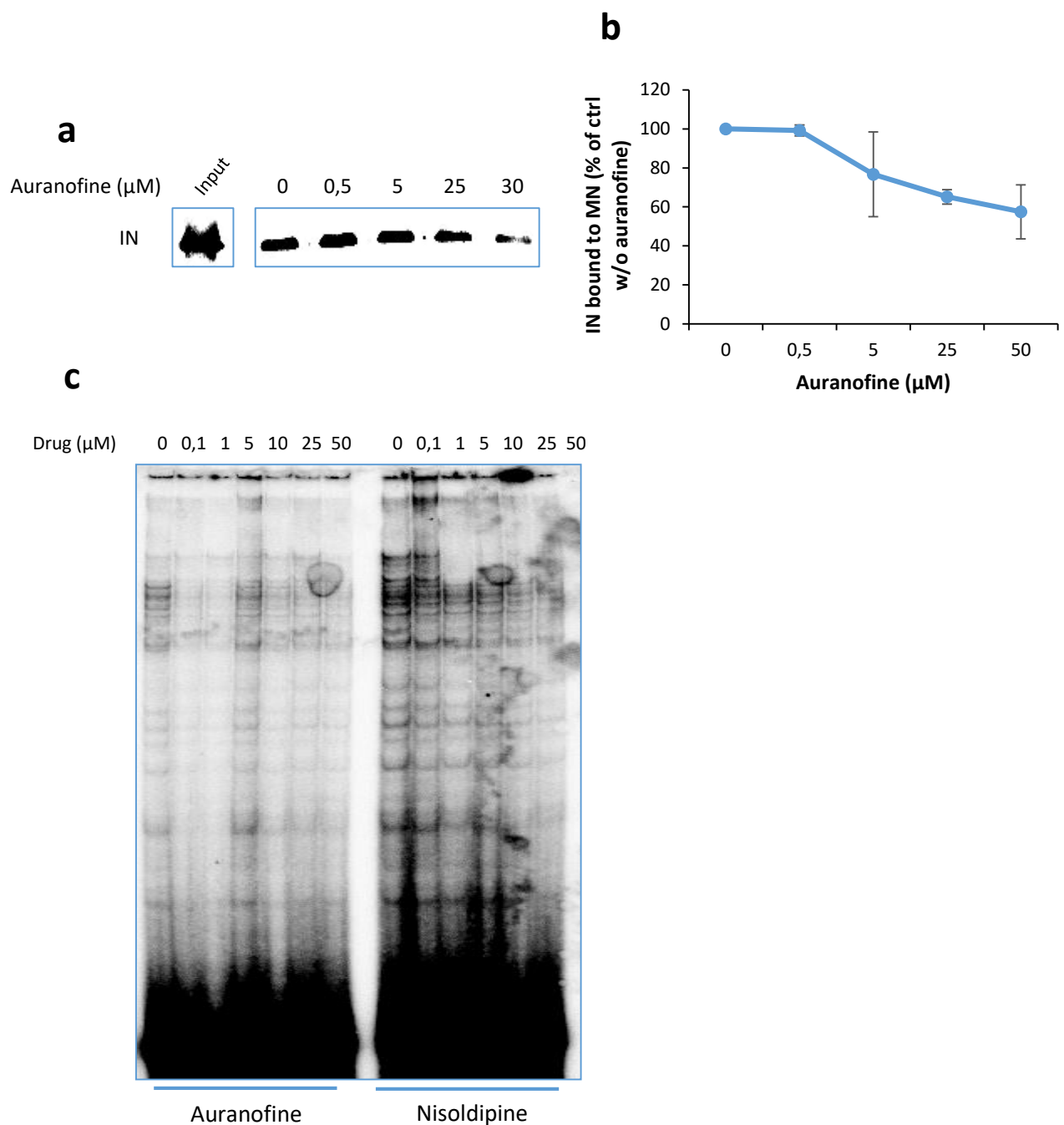


Figure 63 : Effet des molécules alphascreen sur l'interaction et l'intégration *in vitro*. (a) Interaction entre l'IN et le MN en présence d'auranofine par pull down et la quantification en (b). (c) Intégration sur MN en présence d'auranofine (panel de gauche) ou de nisoldipine (panel de droite). Ces résultats ont été réalisés sur deux expériences indépendantes.

rapport à l'inhibition de la réplication virale, il reste encore possible que l'auranofine puisse agir comme inhibiteur de l'intégration en empêchant l'interaction de l'intasome avec le nucléosome.

La deuxième molécule, la nisoldipine, fait partie de la classe des dihydropyridines et est utilisée en tant qu'antihypertenseur. Elle inhibe l'influx transmembranaire des ions calciums dans les muscles cardiaques, ce qui résulte en une dilatation des artérioles faisant ainsi diminuer la pression sanguine. Jusqu'à présent, aucun lien n'a été établi entre la nisoldipine et le VIH. Cependant, des études ont montré que l'infection par le VIH entraîne des dérégulations au niveau des concentrations de calcium intracellulaires qui deviennent anormalement élevées^{187,188}. Un potentiel rôle du calcium lors du cycle viral n'a pas été identifié, mais le fait que le virus puisse influencer sur l'influx de ce dernier est peut-être une piste quant à son implication pour le bon déroulement du cycle viral.

6 Caractérisation biochimique des molécules inhibant l'interaction IN-queue d'histone H4

Les composés ont été testés pour leur capacité à inhiber l'interaction entre l'IN et le nucléosome puis l'intégration sur nucléosome *in vitro*. La caractérisation biochimique des deux composés est encore à l'étude au moment de la rédaction de ce manuscrit, cependant les données préliminaires obtenues sont présentées ici.

Le criblage alphascreen a été réalisé entre le CTD de l'IN et le peptide mimant la queue d'histone H4K20me1, nous avons donc vérifié si les drogues issues de ce criblage pouvaient également inhiber l'interaction entre des partenaires plus physiologiques que sont l'IN entière et le nucléosome. Pour cela les drogues ont été ajoutées dans un pull down réalisé avec ces deux partenaires, et nous avons suivi l'effet des composés sur l'interaction IN-nucléosome. Comme on peut le voir dans la **Figure 63a et b**, l'auranofine est capable de diminuer la quantité d'IN retenue sur les nucléosomes, mais de manière moins efficace que ce que nous avons pu observer en alphascreen (diminution d'environ 40% de l'interaction dans ces conditions). Ceci peut être expliqué par le fait que ce ne sont pas les mêmes partenaires que lors du criblage : nous ne savons pas à l'heure actuelle si l'auranofine se fixe sur l'IN ou au niveau de la queue H4 du nucléosome. Cependant, même si l'effet ne semble pas très efficace dans ces conditions, cela nous conforte tout de même quant à l'utilisation de ces drogues en tant qu'inhibiteurs d'intégration. Nous n'avons pas exploré en détails l'effet de la nisoldipine sur l'interaction IN-nucléosome, mais nos données préliminaires semblent indiquer que cette drogue n'a pas d'effet sur l'interaction.

La seconde étape a été de tester les molécules pour leur capacité à inhiber l'intégration *in vitro* sur nucléosomes. Comme nous l'avons fait pour les mutants, nous avons vérifié dans un premier temps que les molécules n'avaient pas d'effet sur l'intégration *in vitro* sur ADN nu (données non montrées), pour s'assurer que le potentiel effet inhibiteur soit bien lié à la

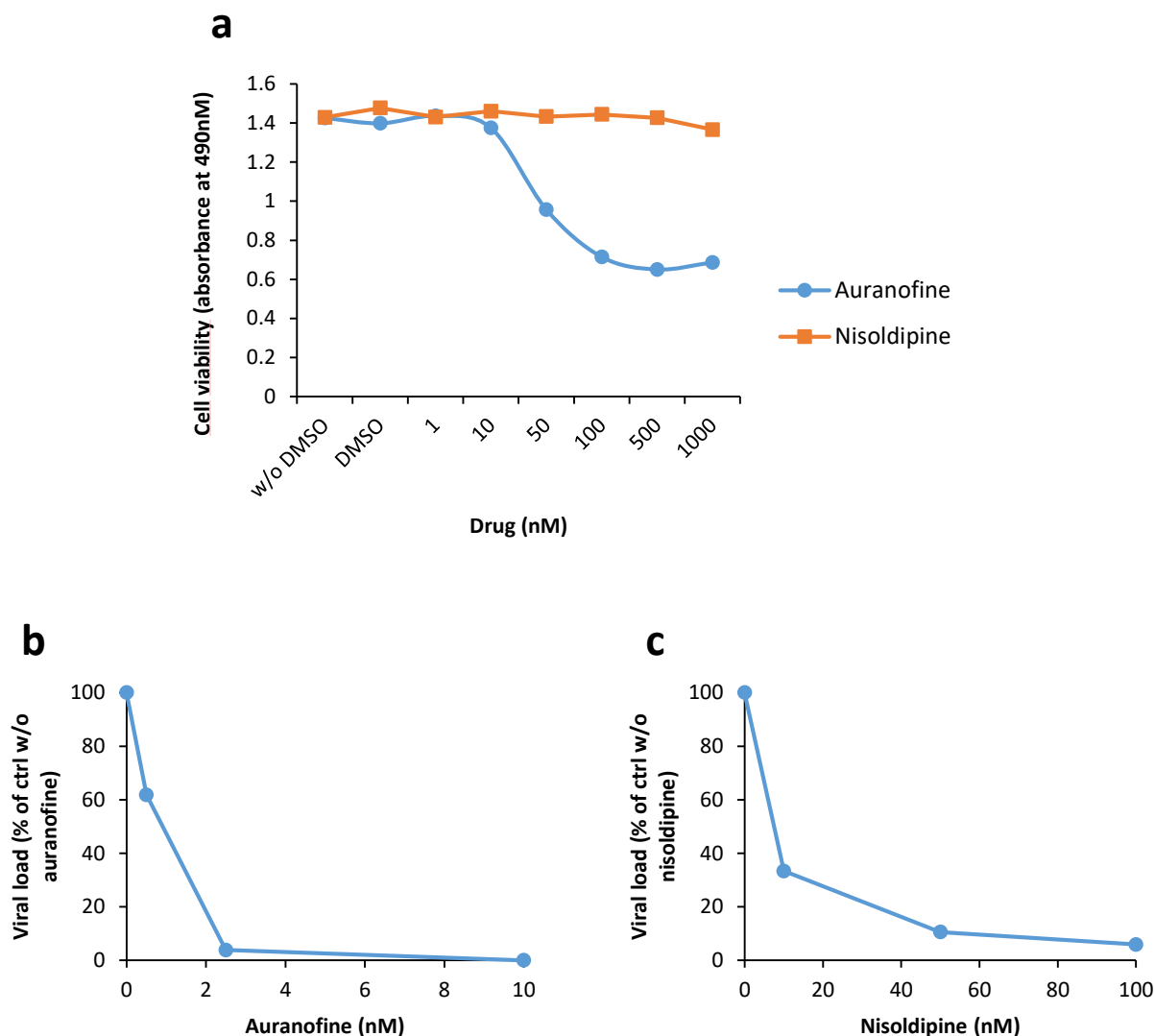


Figure 64 : Effet des molécules alphascreen sur la réplication virale *in vivo*. (a) Analyse de la viabilité des PBMC après 72h de contact avec les molécules alphascreen. **(b)** Analyse de la réplication d'une souche de virus sauvage dans des PBMC en présence d'auranofine ou de nisoldipine **(c)**. Les cellules ont été infectées puis incubées avec le composé choisi. La quantification du nombre de copies d'ARN viral dans le surnageant a été effectuée 72h après l'infection. Ces expériences d'infections ont été en premier lieu réalisées avec des concentrations de drogues plus élevées (de l'ordre de la centaine de nanomolaires) et dans ces conditions nous n'avons pas détecté d'ARN viral dans le surnageant de culture. Les valeurs de charge virale présentées ici sont la résultante d'une seule expérience.

présence des nucléosomes et non un effet sur la catalyse de l'IN. A nouveau, le peu de données que nous avons sur l'effet des drogues sur l'intégration sur nucléosomes ne nous permet pas de conclure quant à un réel effet, cependant il semble y avoir une tendance à l'inhibition pour les deux drogues dans ces conditions (**Figure 63c**).

En résumé, l'effet des deux composés issus du crible alphascreen sur l'interaction et l'intégration *in vitro* sur nucléosomes, il est difficile de conclure au vu du peu de résultats que nous avons obtenus jusqu'à présent. Cependant, même si les effets ne sont pas aussi efficaces que ceux obtenus en alphascreen (avec des IC50 en dessous du μM), il y a tout de même globalement des tendances à l'inhibition.

7 Effet des molécules inhibant l'interaction IN-queue d'histone H4 sur la réplication virale

Afin de choisir le système d'infection le plus adapté pour l'étude des molécules, nous avons vérifié la toxicité de nos deux composés dans différents types cellulaires. Etant donné que les deux molécules sont connues pour avoir des effets anti-cancéreux, nous avons écarté les cellules HeLa et TZM que nous utilisons en routine car ce sont des lignées cancéreuses. C'est la raison pour laquelle nous nous sommes orientés vers un système d'infection plus physiologique basé sur l'infection de PBMC (cellules mononuclées du sang qui sont naturellement infectables par le VIH) et du virus sauvage. Dans ce type cellulaire, l'auranofine présente une toxicité avec une EC50 d'environ 100nM (**Figure 64a**), alors que la nisoldipine ne présente pas de toxicité apparente dans cette gamme de concentration (jusqu'à 1 μM). Basé sur ces données de toxicité, nous avons réalisé les expériences d'infection de PBMC avec une souche de virus sauvage en collaboration avec le Pr H.Fleury et le Dr P.Bellecave (CHU de Bordeaux). Pour ces expériences, nous avons dans un premier temps infecté les PBMC avec le virus avec une multiplicité d'infection de 0,1. 3 heures après les cellules ont été centrifugées et le surnageant a été jeté afin d'éliminer les virus n'ayant pas encore infecté de cellules. Nous avons ensuite ajouté les drogues (à des concentrations non toxiques) en contact avec les cellules infectées. L'analyse de la réplication virale a été effectuée par le CHU 72h après l'infection, en dosant le nombre de copies d'ARN viral dans le surnageant de culture. Les résultats de ces infections ont été très surprenants, car pour les deux drogues nous avons observé une très forte inhibition de la réplication du virus (**Figure 64b et c**) à des concentrations proches du nanomolaire. L'auranofine a une IC50 d'environ 1nM et la nisoldipine d'environ 10nM, ces valeurs semblent indiquer que les molécules sont très efficaces pour inhiber le virus sachant que l'IC50 du dolutégravir (un inhibiteur d'intégration utilisé en thérapie) est aux alentours de 10nM dans ces conditions d'infection (données non montrées). Cependant, bien que ces résultats semblent vraiment prometteurs, il reste à identifier la ou les étapes du cycle viral touchée(s) par les drogues, car nous n'avons pour le moment aucune preuve que ces composés inhibent l'intégration du virus. L'analyse qPCR des différentes populations d'acides nucléiques (ARN et ADN viraux) des cellules infectées, nous

permettra d'identifier l'étape touchée par les drogues, et constitue une perspective directe de ce projet.

8 Conclusion et discussion

L'apparition de virus résistants aux inhibiteurs utilisés en thérapie représente un problème pour le contrôle de l'infection chez certains patients. Le développement de nouveaux composés plus efficaces, mieux tolérés, moins toxiques, est donc important pour pouvoir faire face au plus de cas de figures possibles. Ceci est applicable en particulier pour les inhibiteurs ciblant l'étape d'intégration, cette phase clé du cycle de réplication du virus, car les composés anti-intégration utilisés actuellement en thérapie partagent des mécanismes d'action très proches, ce qui est un inconvénient car le virus développe plus facilement des résistances croisées face à ces composés.

Dans cette étude, nous avons identifié une nouvelle interaction hôte-pathogène qui semble pouvoir servir de base pour l'élaboration de nouvelles stratégies antivirales ciblant l'intégration. En effet par diverses approches nous avons démontré l'importance de l'interaction entre l'intasome et la queue d'histone H4 afin que l'intégration se déroule de manière optimale pour le virus.

Le criblage par alphascreen nous a permis d'identifier des molécules qui sont capables d'empêcher cette interaction *in vitro*, bien que la caractérisation de l'effet de ces molécules est encore en cours, leur effet sur la réplication du virus s'est montré plutôt efficace.

Même si un effet sur l'intégration reste à être démontré, l'ensemble de ces données suggère que l'interaction entre l'intasome du VIH et la queue d'histone H4 peut servir de base pour le développement de nouveaux composés antiviraux ciblant l'intégration. De tels composés seraient très intéressants, car ils disposeraient d'un mécanisme d'action tout à fait différent des INSTIs actuellement utilisés en thérapie, ce qui permettrait de continuer à utiliser des inhibiteurs d'intégration même lorsque le virus aurait développé des résistances aux INSTIs actuels.

Partie 4 : Conclusion générale et perspectives

L'intégration rétrovirale est dépendante de nombreux paramètres, comme illustré par les avancées réalisées ces dernières années sur ce processus. De l'import nucléaire jusqu'au choix final du site d'intégration, de nombreux facteurs ont été identifiés comme influençant l'intégration, et la résolution des structures tri-dimensionnelles d'intasomes de plusieurs rétrovirus différents a permis récemment de mieux comprendre les bases structurales de l'intégration ainsi que les mécanismes de résistance aux INSTIs. Cependant, ces structures soulèvent également plusieurs questions : comment se déroule l'assemblage de l'intasome ? Car la fusion de l'IN du VIH-1 à la protéine Sso7d (qui augmente sa solubilité et son activité) résulte en des intasomes de plusieurs tailles. De plus quels seraient les avantages pour les rétrovirus à posséder des intasomes de tailles différentes ? Une hypothèse est que les intasomes de grosses tailles puissent réaliser plus de contacts avec la chromatine ou même d'être plus sensibles à des signaux associés à la transcription.

Habituellement après l'intégration de l'ADN viral dans le génome de l'hôte, le provirus est transcrit et se réplique. Mais dans certains cas, le génome viral n'est pas transcrit et subsiste dans la chromatine, créant les cellules réservoirs. Des stratégies sont à l'étude pour cibler ces cellules afin d'éliminer les provirus ou de les réactiver pour les faire sortir de leur latence grâce à des composés activateurs de la transcription. Les mécanismes qui influent sur l'établissement d'une intégration latente sont encore peu connus, et certaines questions restent à être élucidées : à quel point l'intégration rétrovirale et la transcription cellulaire sont-elles liées ? Comment le site d'intégration influence l'activité transcriptionnelle du provirus ? Quels sont les déterminants impliqués dans l'établissement d'une intégration latente ? Bien qu'il y ait peu de données pour le moment concernant ces questions, il semble en effet que les sites d'intégrations influencent l'entrée en latence ainsi que la réactivation transcriptionnelle par les LRAs (Latency-Reversing Agents)¹⁸⁹.

Ce travail de thèse a apporté quelques données concernant l'association finale de l'intasome du VIH-1 avec le nucléosome, élucidant certaines de ces questions et ouvrant également de nouvelles voies d'étude.

1 L'interaction IN-queue d'histone H4 comme nouvelle interaction hôte-pathogène

L'influence de la chromatine et plus particulièrement l'accessibilité du nucléosome cible aux intasomes rétroviraux est un sujet étudié depuis des années au laboratoire. Ces travaux ont notamment permis de mettre en évidence que la condensation de la chromatine

n'avait pas la même influence sur l'intégration *in vitro* de différentes IN : HIV et ASV préfèrent des régions pauvres en nucléosomes alors que PFV et MLV au contraire s'intègrent mieux sur des matrices condensées. Cette étude montre que les propriétés intrinsèques des intasomes rétroviraux ainsi que la structure, la densité et la dynamique de la chromatine dans l'environnement du nucléosome ciblé sont des paramètres importants pour la sélectivité d'intégration. De plus, le laboratoire a également identifié que le remodelage de la chromatine était un facteur important pour une intégration optimale *in vitro* et *in vivo*. Ces données suggèrent que des contacts additionnels entre l'intasome et le nucléosome peuvent être modulés par la densité de la chromatine, particulièrement au niveau des queues d'histones dont certaines, comme la queue H4, sont masquées lorsque la chromatine est condensée.

Ceci est soutenu par la découverte d'une interaction directe entre l'IN du VIH-1 et la queue d'histone H4 *in vitro* au laboratoire laissent à penser que cette interaction puisse participer à l'intégration. Nous avons ensuite voulu identifier la zone de l'IN impliquée dans l'interaction avec la queue d'histone H4. Nos études biochimiques ont mené à l'identification d'un site d'interaction aux queues d'histones dans le domaine CTD de l'IN du VIH-1. Ce domaine adopte un repliement proche d'un domaine SH3 connu pour interagir avec des peptides contenant des arginines ou des lysines méthylées, ce qui pourrait corroborer avec une interaction de la queue d'histone H4 méthylée sur la lysine 20, modification interagissant préférentiellement avec l'IN. De plus, la superposition de hMSL3 et du CTD de l'IN montre une très forte homologie structurale. MSL3 ayant la capacité de reconnaître certaines modifications d'histones comme le H4K20me1, il est possible que le CTD de l'IN ait lui aussi une fonction d'interaction voir de lecture de la chromatine.

La suite de l'étude s'est donc focalisée sur ce domaine CTD afin d'étudier plus en détails les bases moléculaires de cette interaction. Nous avons pour cela réalisé une modélisation *in silico* de l'interaction entre le CTD de l'IN et un pentapeptide encadrant la position de la lysine 20 de l'histone H4 (construction utilisée dans la résolution de la structure avec hMSL3). Cette modélisation a révélé une zone d'interaction possible dans le CTD de l'IN au niveau d'une cavité en forme de V pour la fixation du peptide. Dans cette modélisation, le peptide engage 7 liaisons hydrogènes et 15 liaisons hydrophobes avec le CTD de l'IN. De plus, la monométhylation de la lysine 20 du peptide engage à elle seule 7 des 15 liaisons hydrophobes et 1 des 7 liaisons hydrogènes, révélant l'importance de cette monométhylation pour l'interaction. Bien que nous n'ayons pas de structure pour appuyer cette modélisation, nous nous sommes basés sur les interactions prédites par ce modèle afin de générer des mutants d'IN altérés pour leur interaction avec le peptide histone H4. Compte tenu de l'implication du domaine CTD dans diverses étapes du cycle viral (transcription inverse, import nucléaire, interaction avec l'ADN cible, structuration de l'intasome), obtenir des IN mutantes qui ne soient pas affectées pour autre chose que l'interaction avec la queue d'histone H4 fut difficile. Néanmoins, les mutations R231A, R231G, R231H et D253H ont été retenues car elles n'entraînaient pas de perte d'activité catalytique de l'enzyme *in vitro*. Nous avons ensuite évalué l'interaction de ces mutants de l'IN avec la queue d'histone H4 par far dot blot. Ces expériences ont montré que les mutations sur la position R231 entraînaient globalement une baisse d'affinité pour la queue d'histone H4, avec environ 80% de perte d'interaction pour ce

qui concerne le mutant R231H et 40% de perte pour le R231A. Le phénotype opposé a été observé pour le mutant D253H qui présente une meilleure affinité pour la queue d'histone H4 que l'enzyme sauvage (environ 2 fois plus d'interaction). Dans la modélisation, la position R231 du CTD de l'IN est prédite comme interagissant avec le résidu H18 du peptide histone et semble important pour la stabilisation de son extrémité N-terminale. En mutant cette position R231, nous avons obtenu une IN déficiente pour l'interaction avec le peptide, confirmant l'importance pour ce résidu pour l'interaction. La position D253 n'était pas prédite comme interagissant directement avec le peptide, mais semble importante pour la structuration de la cavité en forme de V. En mutant cette position en histidine, il est possible que cela change la structure globale de la cavité, lui faisant adopter une conformation plus affine pour le peptide. Bien que les mutations introduites dans la zone putative de l'interaction avec le peptide affectent clairement l'affinité de l'IN pour la queue d'histone H4 *in vitro*, nous ne pouvons pour le moment pas conclure si ces positions sont impliquées directement ou indirectement dans cette association. Afin d'obtenir plus d'arguments pour valider la modélisation de l'interaction et conforter nos hypothèses de travail, d'autres mutations ont été introduites dans la zone putative d'interaction et sont en cours d'analyse au laboratoire.

Afin de révéler un potentiel rôle de l'interaction IN-queue d'histone H4 lors du cycle viral, nous avons ensuite introduit les mutations de l'IN identifiées précédemment (R231A/G/H et D253H) dans des vecteurs lentiviraux et analysé l'impact de ces mutations sur l'infectivité. Ces expériences ont été réalisées dans plusieurs types cellulaires (293T, HeLa, K562 et TZM) et ont montré que les vecteurs ayant une IN mutée avaient une infectivité altérée par rapport aux vecteurs ayant une IN sauvage. Les vecteurs R231A et R231H présentent une diminution d'infectivité de 40% et 60% respectivement alors que le R231G est plus modéré (baisse d'environ 20%) par rapport au vecteur sauvage, alors que le D253H a une infectivité plus élevée (environ 50% de plus). De manière frappante, ces résultats corrélaient fortement avec la capacité d'interaction des IN avec la queue d'histone H4 *in vitro*, mais pour identifier l'étape du cycle viral qui a été impactée par ces mutations, nous avons quantifié les différentes populations d'ADN virales. Ceci nous a permis d'exclure un défaut de transcription inverse, ce qui aurait pu être attendu étant donnée l'implication de la position R231 pour cette étape. En revanche la quantification de l'ADN intégré a montré une diminution pour les vecteurs R231A et R231H de 30% et 40% respectivement, alors que le phénotype inverse a été observé pour le vecteur D253H avec une augmentation de 30% environ. La quantification des cercles à 2 LTR a permis d'exclure un défaut dans l'import nucléaire. Basés sur ces observations, nous proposons que le défaut d'infectivité (et plus précisément d'intégration) des vecteurs mutants est lié à un défaut d'interaction de l'IN avec la queue d'histone H4 présente sur les nucléosomes.

L'analyse de la distribution des sites d'intégration des cellules infectées par ces vecteurs n'a pas montré de changements drastiques de sélectivité : ils présentent tous une préférence pour les régions transcriptionnellement actives. Cependant, des changements significatifs sont tout de même observés en comparant les mutants entre eux : l'intégration des vecteurs exprimant une IN mutée sur la position R231 a plus tendance à se faire dans les gènes et des régions dynamiques de la chromatine. D'après ces résultats, il semble que

l'interaction IN-queue d'histone H4 soit importante pour l'efficacité d'intégration, mais ne joue pas un rôle très important pour la sélectivité. Etant donné que LEDGF/p75 est un des principaux cofacteurs de l'intégration du VIH en jouant un rôle déterminant pour la sélectivité, il est possible que l'influence de l'interaction IN-H4 sur la sélectivité soit masquée par LEDGF/p75. Pour cela, l'analyse de la distribution des sites d'intégration dans des cellules KO LEDGF/p75 nous permettra de nous affranchir de ce cofacteur afin de voir si l'interaction IN-H4 joue un rôle dans ce contexte. Cette analyse est actuellement en cours au laboratoire.

L'ensemble de ces données supporte l'hypothèse que l'interaction entre l'IN du VIH-1 et la queue d'histone H4 est importante pour le bon déroulement de l'intégration et constitue ainsi une **nouvelle interaction hôte-pathogène**. Afin de confirmer cette hypothèse, nous nous sommes ensuite focalisés sur la caractérisation biochimique de cette interaction afin de comprendre comment elle influence l'intégration.

2 L'activation de l'intasome au contact de la chromatine : une nouvelle étape de l'intégration ?

Au cours de ce projet Maskell et al., ont également identifié une interaction directe entre l'IN de PFV et la queue d'histone H2A grâce à la résolution de la structure du complexe de transfert de brin de PFV. Comme nous l'avons montré pour VIH, cette interaction s'est révélée importante pour l'intégration de PFV *in cellulo*. La caractérisation biochimique de cette interaction leur a permis d'identifier son rôle lors de l'intégration, qui est de participer à l'ancrage de l'intasome sur le nucléosome.

Nous avons observé un phénotype très similaire à leur étude pour ce qui est du cas de l'IN du VIH et de la queue d'histone H4. Des pull down réalisés entre l'IN et des nucléosomes délétés pour une des queues d'histones a révélé un défaut d'interaction de l'IN avec la construction pour laquelle il manque la queue d'histone H4 spécifiquement. De plus, des tests d'intégration réalisés sur ces matrices a également montré que seul ce nucléosome était un moins bon substrat d'intégration. Ces données suggèrent que la queue d'histone H4 est requise pour un ancrage optimal de l'intasome sur le nucléosome *in vitro*.

Pour analyser plus en détails l'effet des queues d'histones sur l'intégration *in vitro*, nous avons utilisé des peptides dérivés de ces différentes queues d'histones que nous avons ajoutés dans des tests d'intégration concertée. Ainsi nous avons pu voir que l'addition du peptide dérivé de la queue d'histone H4 spécifiquement induisait une forte stimulation de l'activité catalytique (3 fois plus d'activité à l'optimum de stimulation). L'utilisation de variants plus ou moins méthylés de la queue d'histone H4 ainsi que les mutants de l'IN altérés pour l'interaction avec ce peptide a permis d'identifier que la stimulation était liée à la capacité d'interaction entre l'IN et le peptide. En effet pour les cas de figures où l'interaction entre ces deux partenaires étaient diminuée (variants de méthylation autres que H4K20me1 ou mutants R231A/H), la stimulation en était également diminuée. Ces données suggèrent qu'une

interaction efficace entre l'IN et l'histone H4 est requise pour une intégration optimale. Cet effet semble spécifique pour l'IN du VIH, car nous n'avons pas observé de stimulation par le peptide H4K20me1 pour les IN de PFV et de RAV. Cependant, l'IN de MVV est aussi stimulée par le peptide, bien que plus faiblement que l'IN du VIH. MVV étant un lentivirus comme VIH, il n'est pas exclu que l'interaction avec l'histone H4 soit spécifique de ce genre de rétrovirus, mais nous n'avons pour le moment pas vérifié l'interaction de l'IN de MVV avec ce peptide.

A ce stade il est difficile de corréliser ces résultats *in vitro* avec un effet physiologique du peptide sur l'intégration. Nous avons donc analysé la nature des produits d'intégration induits par le peptide, et nous avons vu que ces produits étaient ceux qui se rapprochent le plus des produits d'intégration physiologiques (c'est-à-dire de l'intégration full site de deux extrémités de l'ADN viral de manière concertée avec une duplication de paire de bases de 5). Afin de comprendre par quels mécanismes moléculaires le peptide est capable de stimuler l'intégration concertée, nous avons testé le peptide H4 séparément sur les deux réactions que catalyse l'IN : le processing et le transfert de brin. De manière intrigante nous n'avons pas observé de stimulation pour aucune des deux réactions. Nous pensons que ceci est dû à la nature des substrats utilisés dans ces tests qui sont beaucoup plus courts que ceux utilisés dans le test d'intégration concertée. Dans le test de transfert de brin, la réaction d'intégration qui est catalysée est très proche d'une autointégration, car il n'y a qu'un type d'ADN qui sert à la fois de donneur et de receveur. De plus dans le test d'intégration concertée, l'ADN receveur est un plasmide qui présente diverses topologies (super coiled, open circular ...), qui en plus de sa plus grande taille, est peut-être un paramètre impliqué dans la stimulation par le peptide. L'ensemble de ces données suggère que l'effet stimulateur du peptide H4K20me1 soit au niveau de l'étape de transfert de brin. Cette étape est majoritairement dépendante de l'assemblage de l'intasome ainsi que de son association avec l'ADN cible. Afin de discriminer ces deux paramètres, nous avons ensuite évalué l'effet du peptide H4 sur des intasomes déjà assemblés, afin de voir un potentiel effet sur la capture de l'ADN cible et/ou la catalyse de l'intégration.

L'assemblage de l'intasome du VIH-1 représente actuellement une barrière technique à cause de ses propriétés physico-chimiques, les conditions d'assemblage de cet intasome n'ont pour le moment pas été identifiées. C'est la raison pour laquelle nous nous sommes orientés sur l'intasome de MVV, car nous avons vu que cette IN est également stimulée par le peptide H4K20me1 dans le test d'intégration concertée et que les conditions d'assemblage de cet intasome sont décrites. L'ajout du peptide dans des tests d'intégration sur plasmide réalisés avec ces intasomes purifiés a montré que le peptide pouvait également stimuler ces complexes (environ 40% d'activité supplémentaire à l'optimum de stimulation) alors qu'aucun effet n'a été observé sur l'intasome de PFV. Ceci suggère que la stimulation observée ici est reliée à la formation de complexes de transfert de brin (association d'intasomes avec l'ADN cible).

Afin de confirmer ces résultats, nous avons ensuite évalué l'effet du peptide H4K20me1 sur l'interaction entre l'IN et l'ADN cible par pull down. Ces expériences ont montré que le peptide était capable d'augmenter la quantité d'IN retenue sur un ADN 601 (séquence utilisée pour l'assemblage des mononucléosomes) alors que cet effet n'a pas été observé pour

l'ADN HU5 (qui mime un ADN viral). Ces expériences de pull down ont été réalisées avec de l'IN recombinante et non des intasomes, il n'est donc pas exclu que les effets observés ici ne reflètent pas ce qui aurait été obtenu avec des intasomes. En effet au sein d'un intasome certaines régions de l'IN peuvent être masquées à cause des interactions IN-IN, alors qu'avec de l'IN recombinante ces régions seraient accessibles. Il serait donc intéressant de vérifier si des intasomes purifiés ont également plus d'affinité pour de l'ADN cible en présence du peptide (étude en cours au laboratoire).

Bien que nous ayons vu que le peptide pouvait stimuler des intasomes déjà formés, il est également possible que le peptide participe aussi à sa formation. Etant donné qu'un intasome est composé d'un oligomère d'IN, nous avons ensuite évalué l'effet du peptide H4K20me1 sur l'oligomérisation de l'IN par des pontages chimiques au BS³. Ainsi, nous avons pu observer que le peptide avait la capacité d'augmenter la formation d'oligomères de l'IN du VIH, alors qu'aucun effet n'a été observé pour l'IN de PFV. Ces résultats ne permettent pas de conclure quant à un effet du peptide sur la formation d'intasomes, car nous n'avons pas réussi à démontrer que ces espèces oligomériques étaient actives *in vitro*. Cependant il paraît peu probable que ces oligomères soient des agrégats, car si tel était le cas, le peptide ne pourrait pas stimuler l'intégration aussi efficacement.

L'ensemble de ces données (effet sur l'interaction avec l'ADN, effet sur l'oligomérisation, stimulation de l'intégration) suggère que le peptide est capable d'entraîner des changements structuraux au sein de l'IN, et plus particulièrement au niveau de la zone d'interaction du peptide : le CTD. Nous avons donc effectué une étude structurale de RMN sur le CTD de l'IN, en présence ou non du peptide afin de voir si ce dernier pouvait entraîner des changements conformationnels. Dans ces expériences, le spectre du CTD en présence du peptide H4K20me1 présente des différences par rapport au spectre du CTD seul. Ces changements se caractérisent par une déstructuration ainsi que par un effet oligomérisant du peptide. Bien que cette étude ait été conduite sur le CTD seul et non la protéine entière, elle confirme que le peptide histone H4 est capable d'induire des changements structuraux au sein de ce domaine CTD. De plus l'effet oligomérisant observé ici corrèle assez bien avec les expériences de pontages chimiques réalisées précédemment.

Cette deuxième partie de l'étude nous a permis de mieux comprendre les bases moléculaires mises en jeu lors de l'interaction de l'IN avec la queue d'histone H4. L'ensemble de ces résultats indique non seulement que **la queue d'histone H4 est requise pour l'ancrage optimal de l'intasome au nucléosome, mais également que cette interaction induit des changements structuraux dans le CTD de l'IN qui lui permettent d'avoir une meilleure affinité pour l'ADN cible et/ou d'augmenter la formation d'intasomes**. L'implication de ces effets durant le cycle viral reste cependant à être pleinement validée. Cependant, ces données mènent l'équipe à proposer **une nouvelle notion d'activation des intasomes au contact de la chromatine**.

3 L'interaction IN-H4 comme nouvelle cible antivirale

Nos résultats montrent qu'affecter l'association IN-H4 affecte la réplication du virus, indiquant que cette interaction puisse constituer une cible antivirale. La dernière partie de ce projet a ainsi été d'identifier des molécules capables d'empêcher cette interaction. Nous avons employé deux stratégies afin d'identifier de telles molécules : un criblage virtuel basé sur la modélisation moléculaire de l'interaction entre le CTD de l'IN et le pentapeptide H4K20me1 ; et un crible physique par alphascreen. Même si cette dernière partie du projet n'est pas encore finalisée, nous avons tout même pu identifier quelques molécules que nous sommes en train d'étudier au laboratoire.

Le crible virtuel a permis l'identification d'une dizaine de molécules pouvant potentiellement se fixer dans la zone d'interaction identifiée par la modélisation. Le crible de la banque Prestwick par alphascreen a permis d'en identifier deux. Etant donné que les cribles ont été réalisés sur le CTD de l'IN et la queue d'histone H4K20me1, notre stratégie a été de vérifier si les molécules issues du crible étaient capables d'inhiber l'interaction de partenaires plus physiologiques : l'IN entière et le nucléosome. Nous avons commencé la caractérisation de certaines de ces molécules par pull down pour évaluer leur effet sur l'interaction IN-MN, ainsi que dans des tests d'intégration sur MN afin de voir si elles ont la capacité d'inhiber l'intégration *in vitro*. Pour le moment, ces résultats sont très préliminaires et nous nous sommes focalisés sur les deux molécules identifiées en alphascreen : l'auranofine et la nisoldipine. Bien que cela doit être confirmé, nos résultats préliminaires semblent indiquer que ces molécules peuvent inhiber l'intégration sur MN *in vitro*, cependant seule l'auranofine semble capable d'inhiber l'interaction entre l'IN et le MN.

Etant donné que les molécules semblent avoir un effet sur l'intégration *in vitro*, nous les avons testées sur la réplication virale. De manière étonnante, dans un système d'infection de PBMC par une souche de virus sauvage, les deux molécules sont capables d'inhiber la réplication du virus très efficacement à des concentrations de l'ordre du nanomolaire (concentrations de drogues non toxiques pour les cellules). Ces résultats sont très étonnants car il y a une assez grosse différence entre l'efficacité des molécules *in vitro* par rapport aux effets observés en cellules. Cependant, nous ne savons pas pour le moment qu'elle étape du cycle viral a été impactée par les molécules, il est donc difficile d'émettre des hypothèses à ce stade. La quantification des différents acides nucléiques viraux est en cours d'analyse afin d'identifier l'étape touchée.

4 Conclusion générale

Les résultats de cette thèse en complément des données issues du laboratoire et de la littérature de ces dernières années nous ont amené à proposer le modèle d'intégration illustré en **Figure 56**.

Ce modèle fait intervenir différentes notions mais soulève également plusieurs questions :

- L'interaction IN-H4 étant requise pour l'association fonctionnelle de l'intasome au nucléosome suggère que le CTD puisse « scanner » la structure de la chromatine en fonction de l'accessibilité de la queue d'histone H4. Comment cette fonction s'exerce-t-elle dans la cellule et comment est-elle régulée par les facteurs modulant la structure de la chromatine (ex : FACT et SWI/SNF) reste à être identifié.
- La préférence de l'IN pour la forme H4K20me1 *in vitro* et le fait que cette modification soit enrichie au niveau des sites d'intégration du VIH-1, suggère que l'IN ait une fonction de lecture de la chromatine, en faisant la distinction entre différentes modifications d'histones. L'implication de cette marque épigénétique durant le cycle viral reste cependant à être confirmée.
- La stimulation de l'intégration par la queue d'histone H4 semble être dépendante de plusieurs paramètres tels que des changements structuraux dans le CTD de l'IN entraînant possiblement une oligomérisation de l'enzyme ainsi qu'une augmentation de l'affinité pour l'ADN. La relevance biologique de cette étape d'activation de l'intasome au contact de la queue d'histone H4 reste cependant à être confirmée dans le contexte cellulaire.
- Récemment, il a été mis en évidence que l'ADN viral de MLV était chromatinisé avant d'entrer en contact avec la chromatine¹⁹⁰. Ce mécanisme n'a pour le moment pas été identifié pour HIV, mais cela soulève la possibilité que les histones puissent entrer en contact avec les complexes de pré-intégration. Etant donné les effets que nous avons observé résultant de l'interaction entre l'IN et la queue d'histone H4, il n'est pas exclu qu'un contact précoce entre ces deux partenaires puisse influencer sur l'intégration du VIH.

BIBLIOGRAPHIE

1. Barré-Sinoussi, F. *et al.* Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science, New Series* **220**, 868–871 (1983).
2. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. Detection, Isolation, and Continuous Production of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS and Pre-AIDS. *Science, New Series* **224**, 497–500 (1984).
3. Skalka, A. M. Retroviral DNA Transposition: Themes and Variations. **22** (2015).
4. Campbell, E. M. & Hope, T. J. HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nature Reviews Microbiology* **13**, 471–483 (2015).
5. Hemelaar, J. The origin and diversity of the HIV-1 pandemic. *Trends Mol Med* **18**, 182–192 (2012).
6. Costin, J. M. Cytopathic Mechanisms of HIV-1. *Virology Journal* **4**, 100 (2007).
7. Neurath, A. R., Strick, N., Jiang, S., Li, Y.-Y. & Debnath, A. K. Anti-HIV-1 activity of cellulose acetate phthalate: Synergy with soluble CD4 and induction of ‘dead-end’ gp41 six-helix bundles. *BMC Infectious Diseases* **2**, (2002).
8. Pluta, K. & Kacprzak, M. M. Use of HIV as a gene transfer vector. *Lentiviral vectors* **56**, 65 (2009).
9. Blumenthal, R., Durell, S. & Viard, M. HIV Entry and Envelope Glycoprotein-mediated Fusion. *J Biol Chem* **287**, 40841–40849 (2012).
10. Buzon, V. *et al.* Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathog.* **6**, e1000880 (2010).
11. Bukrinsky, M. I. *et al.* A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature* **365**, 666–669 (1993).
12. Bukrinskaya, A. HIV-1 matrix protein: a mysterious regulator of the viral life cycle. *Virus Res.* **124**, 1–11 (2007).
13. Pornillos, O., Ganser-Pornillos, B. K. & Yeager, M. Atomic-level modelling of the HIV capsid. *Nature* **469**, 424–427 (2011).
14. Fassati, A. Multiple roles of the capsid protein in the early steps of HIV-1 infection. *Virus Res.* **170**, 15–24 (2012).
15. Matreyek, K. A. & Engelman, A. Viral and Cellular Requirements for the Nuclear Entry of Retroviral Preintegration Nucleoprotein Complexes. *Viruses* **5**, 2483–2511 (2013).
16. Sowd, G. A. *et al.* 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 (2016).
17. Brass, A. L. *et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319**, 921–926 (2008).

18. Matreyek, K. A., Yücel, S. S., Li, X. & Engelman, A. Nucleoporin NUP153 Phenylalanine-Glycine Motifs Engage a Common Binding Pocket within the HIV-1 Capsid Protein to Mediate Lentiviral Infectivity. *PLOS Pathogens* **9**, e1003693 (2013).
19. Schaller, T. *et al.* HIV-1 Capsid-Cyclophilin Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency. *PLOS Pathogens* **7**, e1002439 (2011).
20. Ilina, T., LaBarge, K., Sarafianos, S. G., Ishima, R. & Parniak, M. A. Inhibitors of HIV-1 Reverse Transcriptase—Associated Ribonuclease H Activity. *Biology* **1**, 521–541 (2012).
21. De Guzman, R. N. *et al.* Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **279**, 384–388 (1998).
22. Levin, J. G., Mitra, M., Mascarenhas, A. & Musier-Forsyth, K. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biology* **7**, 754–774 (2010).
23. Friedrich, M. *et al.* Glutamic Acid Residues in HIV-1 p6 Regulate Virus Budding and Membrane Association of Gag. *Viruses* **8**, 117 (2016).
24. Navia, M. A. *et al.* Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **337**, 615–620 (1989).
25. Yang, H., Nkeze, J. & Zhao, R. Y. Effects of HIV-1 protease on cellular functions and their potential applications in antiretroviral therapy. *Cell Biosci* **2**, 32 (2012).
26. Jacobo-Molina, A. *et al.* Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6320–6324 (1993).
27. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783–1790 (1992).
28. Romani, B., Engelbrecht, S. & Glashoff, R. H. Functions of Tat: the versatile protein of human immunodeficiency virus type 1. *J. Gen. Virol.* **91**, 1–12 (2010).
29. Musinova, Y. R., Sheval, E. V., Dib, C., Germini, D. & Vassetzky, Y. S. Functional roles of HIV-1 Tat protein in the nucleus. *Cell. Mol. Life Sci.* **73**, 589–601 (2016).
30. Groom, H. C. T., Anderson, E. C., Dangerfield, J. A. & Lever, A. M. L. Rev regulates translation of human immunodeficiency virus type 1 RNAs. *J. Gen. Virol.* **90**, 1141–1147 (2009).
31. Margottin, F. *et al.* A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. *Mol. Cell* **1**, 565–574 (1998).
32. Li, L., Li, H. S., Pauza, C. D., Bukrinsky, M. & Zhao, R. Y. Roles of HIV-1 auxiliary proteins in viral pathogenesis and host-pathogen interactions. *Cell Res.* **15**, 923–934 (2005).

33. Andrew, A. & Strebel, K. HIV-1 Vpu targets cell surface markers CD4 and BST-2 through distinct mechanisms. *Mol Aspects Med* **31**, 407–417 (2010).
34. Engelman, A. & Cherepanov, P. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature Reviews Microbiology* **10**, 279–290 (2012).
35. Kim, K. *et al.* Vpr-Binding Protein Antagonizes p53-Mediated Transcription via Direct Interaction with H3 Tail. *Mol Cell Biol* **32**, 783–796 (2012).
36. Connor, R. I., Chen, B. K., Choe, S. & Landau, N. R. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. *Virology* **206**, 935–944 (1995).
37. Marin, M., Rose, K. M., Kozak, S. L. & Kabat, D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat. Med.* **9**, 1398–1403 (2003).
38. Maddon, P. J. *et al.* The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**, 333–348 (1986).
39. Wilen, C. B., Tilton, J. C. & Doms, R. W. HIV: Cell Binding and Entry. *Cold Spring Harb Perspect Med* **2**, (2012).
40. Jacquot, G. & Benichou, S. Import nucléaire du matériel génétique du virus de l'immunodéficience humaine de type. **10**, 11 (2006).
41. Lusic, M. & Siliciano, R. F. Nuclear landscape of HIV-1 infection and integration. *Nature Reviews Microbiology* **15**, 69–82 (2017).
42. Lori, F. *et al.* Viral DNA carried by human immunodeficiency virus type 1 virions. *J. Virol.* **66**, 5067–5074 (1992).
43. Trono, D. Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses. *J. Virol.* **66**, 4893–4900 (1992).
44. Zhu, J. & Cunningham, J. M. Minus-strand DNA is present within murine type C ecotropic retroviruses prior to infection. *J. Virol.* **67**, 2385–2388 (1993).
45. Hu, W.-S. & Hughes, S. H. HIV-1 Reverse Transcription. *Cold Spring Harb Perspect Med* **2**, (2012).
46. Farnet, C. M. & Haseltine, W. A. Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J Virol* **65**, 1910–1915 (1991).
47. Yamashita, M. & Emerman, M. Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *J. Virol.* **78**, 5670–5678 (2004).
48. Koh, Y. *et al.* 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. *J. Virol.* **87**, 648–658 (2013).

49. Di Nunzio, F. *et al.* Nup153 and Nup98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. *Virology* **440**, 8–18 (2013).
50. Lee, K. *et al.* Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe* **7**, 221–233 (2010).
51. Bhattacharya, A. *et al.* Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18625–18630 (2014).
52. Rasheedi, S. *et al.* The Cleavage and Polyadenylation Specificity Factor 6 (CPSF6) Subunit of the Capsid-recruited Pre-messenger RNA Cleavage Factor I (CFIm) Complex Mediates HIV-1 Integration into Genes. *J. Biol. Chem.* **291**, 11809–11819 (2016).
53. Chin, C. R. *et al.* Direct Visualization of HIV-1 Replication Intermediates Shows that Capsid and CPSF6 Modulate HIV-1 Intra-nuclear Invasion and Integration. *Cell Rep* **13**, 1717–1731 (2015).
54. Gamble, T. R. *et al.* Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87**, 1285–1294 (1996).
55. Bosco, D. A., Eisenmesser, E. Z., Pochapsky, S., Sundquist, W. I. & Kern, D. Catalysis of cis/trans isomerization in native HIV-1 capsid by human cyclophilin A. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5247–5252 (2002).
56. Jones, K. A. & Peterlin, B. M. Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.* **63**, 717–743 (1994).
57. Karn, J. & Stoltzfus, C. M. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med* **2**, a006916 (2012).
58. Sundquist, W. I. & Krausslich, H.-G. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harbor Perspectives in Medicine* **2**, a006924–a006924 (2012).
59. Butsch, M. & Boris-Lawrie, K. Destiny of unspliced retroviral RNA: ribosome and/or virion? *J. Virol.* **76**, 3089–3094 (2002).
60. Adamson, C. S. & Freed, E. O. Human immunodeficiency virus type 1 assembly, release, and maturation. *Adv. Pharmacol.* **55**, 347–387 (2007).
61. Demirov, D. G. & Freed, E. O. Retrovirus budding. *Virus Res.* **106**, 87–102 (2004).
62. Tritch, R. J., Cheng, Y. E., Yin, F. H. & Erickson-Viitanen, S. Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 gag polyprotein. *J. Virol.* **65**, 922–930 (1991).
63. Pettit, S. C., Everitt, L. E., Choudhury, S., Dunn, B. M. & Kaplan, A. H. Initial Cleavage of the Human Immunodeficiency Virus Type 1 GagPol Precursor by Its Activated Protease Occurs by an Intramolecular Mechanism. *J Virol* **78**, 8477–8485 (2004).
64. Molina, J.-M. *et al.* On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. *N. Engl. J. Med.* **373**, 2237–2246 (2015).

65. Thorlund, K., Horwitz, M. S., Fife, B. T., Lester, R. & Cameron, D. W. Landscape review of current HIV 'kick and kill' cure research - some kicking, not enough killing. *BMC Infectious Diseases* **17**, (2017).
66. Descours, B. *et al.* CD32a is a marker of a CD4 T-cell HIV reservoir harbouring replication-competent proviruses. *Nature* **543**, 564–567 (2017).
67. Zhang, D., Li, W. & Jiang, S. Peptide fusion inhibitors targeting the HIV-1 gp41: a patent review (2009 - 2014). *Expert Opin Ther Pat* **25**, 159–173 (2015).
68. Xu, F. *et al.* Current Status of the Pharmacokinetics and Pharmacodynamics of HIV-1 Entry Inhibitors and HIV Therapy. *Curr. Drug Metab.* **18**, 769–781 (2017).
69. Lesbats, P., Engelman, A. N. & Cherepanov, P. Retroviral DNA Integration. *Chemical Reviews* **116**, 12730–12757 (2016).
70. Tressler, R. & Godfrey, C. NRTI backbone in HIV treatment: will it remain relevant? *Drugs* **72**, 2051–2062 (2012).
71. de Béthune, M.-P. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989-2009). *Antiviral Res.* **85**, 75–90 (2010).
72. Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cherepanov, P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236 (2010).
73. Koh, Y., Matreyek, K. A. & Engelman, A. Differential sensitivities of retroviruses to integrase strand transfer inhibitors. *J. Virol.* **85**, 3677–3682 (2011).
74. Christ, F. *et al.* Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nat. Chem. Biol.* **6**, 442–448 (2010).
75. Tsiang, M. *et al.* New class of HIV-1 integrase (IN) inhibitors with a dual mode of action. *J. Biol. Chem.* **287**, 21189–21203 (2012).
76. Kessl, J. J. *et al.* Multimode, cooperative mechanism of action of allosteric HIV-1 integrase inhibitors. *J. Biol. Chem.* **287**, 16801–16811 (2012).
77. Christ, F. *et al.* Small-molecule inhibitors of the LEDGF/p75 binding site of integrase block HIV replication and modulate integrase multimerization. *Antimicrob. Agents Chemother.* **56**, 4365–4374 (2012).
78. Le Rouzic, E. *et al.* Dual inhibition of HIV-1 replication by integrase-LEDGF allosteric inhibitors is predominant at the post-integration stage. *Retrovirology* **10**, 144 (2013).
79. Kessl, J. J. *et al.* HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell* **166**, 1257-1268.e12 (2016).
80. Lv, Z., Chu, Y. & Wang, Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl)* **7**, 95–104 (2015).

81. Rerks-Ngarm, S. *et al.* Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N. Engl. J. Med.* **361**, 2209–2220 (2009).
82. Engelman, A. & Craigie, R. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J. Virol.* **66**, 6361–6369 (1992).
83. Cai, M. *et al.* Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* **4**, 567–577 (1997).
84. Dyda, F. *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986 (1994).
85. Jenkins, T. M., Engelman, A., Ghirlando, R. & Craigie, R. A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J. Biol. Chem.* **271**, 7712–7718 (1996).
86. Lodi, P. J. *et al.* Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* **34**, 9826–9833 (1995).
87. Engelman, A. N. & Cherepanov, P. Retroviral intasomes arising. *Current Opinion in Structural Biology* **47**, 23–29 (2017).
88. Yin, Z. *et al.* Crystal structure of the Rous sarcoma virus intasome. *Nature* **530**, 362–366 (2016).
89. Ballandras-Colas, A. *et al.* Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* **530**, 358–361 (2016).
90. Ballandras-Colas, A. *et al.* A supramolecular assembly mediates lentiviral DNA integration. *Science* **355**, 93–95 (2017).
91. Li, M., Jurado, K. A., Lin, S., Engelman, A. & Craigie, R. Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLoS ONE* **9**, e105078 (2014).
92. Passos, D. O. *et al.* CryoEM Structures and Atomic Model of the HIV-1 Strand Transfer Complex Intasome. *Science* **355**, 89–92 (2017).
93. Miller, M. D., Farnet, C. M. & Bushman, F. D. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* **71**, 5382–5390 (1997).
94. Sherman, P. A. & Fyfe, J. A. Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5119–5123 (1990).
95. Thierry, E., Deprez, E. & Delelis, O. Different Pathways Leading to Integrase Inhibitors Resistance. *Frontiers in Microbiology* **7**, (2017).
96. Yoder, K. E. & Bushman, F. D. Repair of gaps in retroviral DNA integration intermediates. *J. Virol.* **74**, 11191–11200 (2000).

97. Cosnefroy, O. *et al.* Stimulation of the human RAD51 nucleofilament restricts HIV-1 integration in vitro and in infected cells. *J. Virol.* **86**, 513–526 (2012).
98. Bukrinsky, M. I. *et al.* Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6580–6584 (1992).
99. Semenova, E. A. Preferential Inhibition of the Magnesium-Dependent Strand Transfer Reaction of HIV-1 Integrase by -Hydroxytropolones. *Molecular Pharmacology* **69**, 1454–1460 (2006).
100. Chow, S. A., Vincent, K. A., Ellison, V. & Brown, P. O. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* **255**, 723–726 (1992).
101. Engelman, A. In vivo analysis of retroviral integrase structure and function. *Adv. Virus Res.* **52**, 411–426 (1999).
102. Engelman, A. Pleiotropic Nature of HIV-1 Integrase Mutations. in *HIV-1 Integrase* 67–81 (Wiley-Blackwell, 2011). doi:10.1002/9781118015377.ch6
103. Tekeste, S. S. *et al.* Interaction between Reverse Transcriptase and Integrase Is Required for Reverse Transcription during HIV-1 Replication. *Journal of Virology* **89**, 12058–12069 (2015).
104. Bouyac-Bertoia, M. *et al.* HIV-1 Infection Requires a Functional Integrase NLS. *Molecular Cell* **7**, 1025–1035 (2001).
105. De Houwer, S. *et al.* Identification of Residues in the C-terminal Domain of HIV-1 Integrase That Mediate Binding to the Transportin-SR2 Protein. *J Biol Chem* **287**, 34059–34068 (2012).
106. Kalpana, G. V., Marmon, S., Wang, W., Crabtree, G. R. & Goff, S. P. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* **266**, 2002–2006 (1994).
107. Lesbats, P. *et al.* Functional coupling between HIV-1 integrase and the SWI/SNF chromatin remodeling complex for efficient in vitro integration into stable nucleosomes. *PLoS Pathog.* **7**, e1001280 (2011).
108. Yung, E. *et al.* Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1. *Nat. Med.* **7**, 920–926 (2001).
109. Engelman, A. N. & Singh, P. K. Cellular and molecular mechanisms of HIV-1 integration targeting. *Cellular and Molecular Life Sciences* **75**, 2491–2507 (2018).
110. Lin, C.-W. & Engelman, A. The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. *J. Virol.* **77**, 5030–5036 (2003).

111. Li, L. *et al.* Retroviral cDNA Integration: Stimulation by HMG I Family Proteins. *J Virol* **74**, 10965–10974 (2000).
112. Henderson, A., Holloway, A., Reeves, R. & Tremethick, D. J. Recruitment of SWI/SNF to the human immunodeficiency virus type 1 promoter. *Mol. Cell. Biol.* **24**, 389–397 (2004).
113. Orphanides, G., Wu, W. H., Lane, W. S., Hampsey, M. & Reinberg, D. The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* **400**, 284–288 (1999).
114. Hsieh, F.-K. *et al.* Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7654–7659 (2013).
115. Matysiak, J. *et al.* Modulation of chromatin structure by the FACT histone chaperone complex regulates HIV-1 integration. *Retrovirology* **14**, 39 (2017).
116. Winans, S. *et al.* The FACT Complex Promotes Avian Leukosis Virus DNA Integration. *Journal of Virology* **91**, (2017).
117. Debyser, Z., Christ, F., De Rijck, J. & Gijssbers, R. Host factors for retroviral integration site selection. *Trends in Biochemical Sciences* **40**, 108–116 (2015).
118. Daugaard, M. *et al.* LEDGF (p75) promotes DNA-end resection and homologous recombination. *Nat. Struct. Mol. Biol.* **19**, 803–810 (2012).
119. Cherepanov, P., Devroe, E., Silver, P. A. & Engelman, A. 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**, 48883–48892 (2004).
120. Eidahl, J. O. *et al.* Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Res.* **41**, 3924–3936 (2013).
121. van Nuland, R. *et al.* Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics Chromatin* **6**, 12 (2013).
122. Busschots, K. *et al.* The Interaction of LEDGF/p75 with Integrase Is Lentivirus-specific and Promotes DNA Binding. *Journal of Biological Chemistry* **280**, 17841–17847 (2005).
123. Cherepanov, P. LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res.* **35**, 113–124 (2007).
124. Donnelly, N., Gorman, A. M., Gupta, S. & Samali, A. The eIF2 α kinases: their structures and functions. *Cell. Mol. Life Sci.* **70**, 3493–3511 (2013).
125. Jaspart, A. *et al.* GCN2 phosphorylates HIV-1 integrase and decreases HIV-1 replication by limiting viral integration. *Sci Rep* **7**, (2017).
126. Galkin, V. E. *et al.* The Rad51/RadA N-terminal domain activates nucleoprotein filament ATPase activity. *Structure* **14**, 983–992 (2006).

127. Maskell, D. P. *et al.* Structural basis for retroviral integration into nucleosomes. *Nature* **523**, 366–369 (2015).
128. Holman, A. G. & Coffin, J. M. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6103–6107 (2005).
129. Wu, X., Li, Y., Crise, B., Burgess, S. M. & Munroe, D. J. Weak Palindromic Consensus Sequences Are a Common Feature Found at the Integration Target Sites of Many Retroviruses. *J Virol* **79**, 5211–5214 (2005).
130. Luger, K. Crystal structure of the nucleosome core particle at 2.8 Å resolution. **389**, 10 (1997).
131. Arents, G. & Moudrianakis, E. N. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11170–11174 (1995).
132. Tessarz, P. *et al.* Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* **505**, 564–568 (2014).
133. Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Biol.* **6**, 838–849 (2005).
134. Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**, 2343–2360 (2001).
135. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
136. Black, J. C., Van Rechem, C. & Whetstone, J. R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell* **48**, 491–507 (2012).
137. Maurer-Stroh, S. *et al.* The Tudor domain ‘Royal Family’: Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci.* **28**, 69–74 (2003).
138. Peterson, C. L. & Laniel, M.-A. Histones and histone modifications. *Curr. Biol.* **14**, R546–551 (2004).
139. Choi, J. K. & Howe, L. J. Histone acetylation: truth of consequences? *Biochem. Cell Biol.* **87**, 139–150 (2009).
140. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
141. van Attikum, H., Fritsch, O. & Gasser, S. M. Distinct roles for SWR1 and INO80 chromatin remodeling complexes at chromosomal double-strand breaks. *EMBO J.* **26**, 4113–4125 (2007).

142. Fink, M., Imholz, D. & Thoma, F. Contribution of the serine 129 of histone H2A to chromatin structure. *Mol. Cell. Biol.* **27**, 3589–3600 (2007).
143. Cao, J. & Yan, Q. Histone Ubiquitination and Deubiquitination in Transcription, DNA Damage Response, and Cancer. *Frontiers in Oncology* **2**, (2012).
144. Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78**, 273–304 (2009).
145. Becker, P. B. & Workman, J. L. Nucleosome remodeling and epigenetics. *Cold Spring Harb Perspect Biol* **5**, (2013).
146. Mueller-Planitz, F., Klinker, H. & Becker, P. B. Nucleosome sliding mechanisms: new twists in a looped history. *Nat. Struct. Mol. Biol.* **20**, 1026–1032 (2013).
147. Kireeva, M. L. *et al.* Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Mol. Cell* **9**, 541–552 (2002).
148. Desfarges, S. & Ciuffi, A. Retroviral Integration Site Selection. *Viruses* **2**, 111–130 (2010).
149. Schröder, A. R. W. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521–529 (2002).
150. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
151. Mitchell, R. S. *et al.* Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2**, E234 (2004).
152. Ocwieja, K. E. *et al.* HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS Pathog.* **7**, e1001313 (2011).
153. Saito, A. *et al.* Capsid-CPSF6 Interaction Is Dispensable for HIV-1 Replication in Primary Cells but Is Selected during Virus Passage In Vivo. *J. Virol.* **90**, 6918–6935 (2016).
154. Marini, B. *et al.* Nuclear architecture dictates HIV-1 integration site selection. *Nature* **521**, 227–231 (2015).
155. Albanese, A., Arosio, D., Terreni, M. & Cereseto, A. HIV-1 Pre-Integration Complexes Selectively Target Decondensed Chromatin in the Nuclear Periphery. *PLoS One* **3**, (2008).
156. Di Nunzio, F. *et al.* Human Nucleoporins Promote HIV-1 Docking at the Nuclear Pore, Nuclear Import and Integration. *PLoS One* **7**, (2012).
157. Lelek, M. *et al.* Chromatin organization at the nuclear pore favours HIV replication. *Nat Commun* **6**, (2015).
158. Wong, R. W., Mamede, J. I. & Hope, T. J. Impact of Nucleoporin-Mediated Chromatin Localization and Nuclear Architecture on HIV Integration Site Selection. *J Virol* **89**, 9702–9705 (2015).

159. Kvaratskhelia, M., Sharma, A., Larue, R. C., Serrao, E. & Engelman, A. Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Research* **42**, 10209–10225 (2014).
160. Wu, X., Li, Y., Crise, B. & Burgess, S. M. Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**, 1749–1751 (2003).
161. LaFave, M. C. *et al.* MLV integration site selection is driven by strong enhancers and active promoters. *Nucleic Acids Res* **42**, 4257–4269 (2014).
162. De Ravin, S. S. *et al.* Enhancers are major targets for murine leukemia virus vector integration. *J. Virol.* **88**, 4504–4513 (2014).
163. Sharma, A. *et al.* BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12036–12041 (2013).
164. Gupta, S. S. *et al.* Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J. Virol.* **87**, 12721–12736 (2013).
165. De Rijck, J. *et al.* The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep* **5**, 886–894 (2013).
166. Larue, R. C. *et al.* Bimodal high-affinity association of Brd4 with murine leukemia virus integrase and mononucleosomes. *Nucleic Acids Res.* **42**, 4868–4881 (2014).
167. Aiyer, S. *et al.* Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction. *Nucleic Acids Res.* **42**, 5917–5928 (2014).
168. Wu, S.-Y. & Chiang, C.-M. The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *J. Biol. Chem.* **282**, 13141–13145 (2007).
169. Chiang, C.-M. Brd4 engagement from chromatin targeting to transcriptional regulation: selective contact with acetylated histone H3 and H4. *F1000 Biol Rep* **1**, 98 (2009).
170. Umehara, T. *et al.* Structural basis for acetylated histone H4 recognition by the human BRD2 bromodomain. *J. Biol. Chem.* **285**, 7610–7618 (2010).
171. Cherepanov, P., Ambrosio, A. L. B., Rahman, S., Ellenberger, T. & Engelman, A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17308–17313 (2005).
172. Maertens, G. *et al.* LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J. Biol. Chem.* **278**, 33528–33539 (2003).
173. Vandegraaff, N., Devroe, E., Turlure, F., Silver, P. A. & Engelman, A. Biochemical and genetic analyses of integrase-interacting proteins lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication. *Virology* **346**, 415–426 (2006).

174. Wang, H. *et al.* 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 Research* **40**, 11518–11530 (2012).
175. Schrijvers, R. *et al.* HRP-2 determines HIV-1 integration site selection in LEDGF/p75 depleted cells. *Retrovirology* **9**, 84 (2012).
176. Serrao, E., Ballandras-Colas, A., Cherepanov, P., Maertens, G. N. & Engelman, A. N. Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology* **12**, (2015).
177. Benleulmi, M. *et al.* Intasome architecture and chromatin density modulate retroviral integration into nucleosome. *Retrovirology* **12**, 13 (2015).
178. Benleulmi, M. S. *et al.* Modulation of the functional association between the HIV-1 intasome and the nucleosome by histone amino-terminal tails. *Retrovirology* **14**, (2017).
179. Brussel, A. & Sonigo, P. Analysis of Early Human Immunodeficiency Virus Type 1 DNA Synthesis by Use of a New Sensitive Assay for Quantifying Integrated Provirus. *Journal of Virology* **77**, 10119–10124 (2003).
180. Eijkelenboom, A. P. *et al.* The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat. Struct. Biol.* **2**, 807–810 (1995).
181. Kim, D. *et al.* Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nature Structural & Molecular Biology* **17**, 1027–1029 (2010).
182. Moore, S. A., Ferhatoglu, Y., Jia, Y., Al-Jiab, R. A. & Scott, M. J. Structural and biochemical studies on the chromo-barrel domain of male specific lethal 3 (MSL3) reveal a binding preference for mono- or dimethyllysine 20 on histone H4. *J. Biol. Chem.* **285**, 40879–40890 (2010).
183. Faure, A. *et al.* HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res.* **33**, 977–986 (2005).
184. Chirullo, B. *et al.* A candidate anti-HIV reservoir compound, auranofin, exerts a selective ‘anti-memory’ effect by exploiting the baseline oxidative status of lymphocytes. *Cell Death & Disease* **4**, e944–e944 (2013).
185. Reiser, K. *et al.* Thioredoxin-1 and protein disulfide isomerase catalyze the reduction of similar disulfides in HIV gp120. *The International Journal of Biochemistry & Cell Biology* **44**, 556–562 (2012).
186. Reiser, K. *et al.* The Cellular Thioredoxin-1/Thioredoxin Reductase-1 Driven Oxidoreduction Represents a Chemotherapeutic Target for HIV-1 Entry Inhibition. *PLOS ONE* **11**, e0147773 (2016).
187. Bennasser, Y. *et al.* [HIV-1 Tat protein induces IL-10 production by human monocytes: implications of the PKC and calcium pathway]. *J. Soc. Biol.* **195**, 319–326 (2001).

188. Hegg, C. C., Hu, S., Peterson, P. K. & Thayer, S. A. Beta-chemokines and human immunodeficiency virus type-1 proteins evoke intracellular calcium increases in human microglia. *Neuroscience* **98**, 191–199 (2000).
189. Chen, H.-C., Martinez, J. P., Zorita, E., Meyerhans, A. & Fillion, G. J. Position effects influence HIV latency reversal. *Nat. Struct. Mol. Biol.* **24**, 47–54 (2017).
190. Wang, G. Z., Wang, Y. & Goff, S. P. Histones Are Rapidly Loaded onto Unintegrated Retroviral DNAs Soon after Nuclear Entry. *Cell Host & Microbe* **20**, 798–809 (2016).