UNIVERSITÉ DE STRASBOURG

Ecole doctorale des sciences de la vie et de la santé

CNRS, UPR9002 : ARN

THÈSE

présentée par : **Marine KANJA**

soutenue le : 28 septembre 2017

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Science du vivant

Spécialité : Aspects moléculaires et cellulaires de la biologie

Coévolution dans le gène *pol* du VIH-1 : un carrefour aux frontières de nouvelles espèces du VIH

THÈSE dirigée par :

NEGRONI Matteo — Directeur de recherche (CNRS), Université de Strasbourg

RAPPORTEURS EXTERNES :

GOUET Patrice — Professeur, Université de Lyon

LAVIGNE Marc — Chargé de recherche (CNRS), Institut Pasteur, Paris

---

EXAMINATEUR INTERNE :

DIMITROVA Maria — Maître de conférence, Université de Strasbourg

MEMBRE INVITE – ENCADRANT de la thèse :

LENER Daniela — Maître de conférence, Université de Strasbourg

Sidaction
ENSEMBLE CONTRE LE SIDA

Remerciements

Je tenais tout d'abord à remercier les membres de jury, Patrice Gouet, Marc Lavigne et Maria Dimitrova, pour avoir accepter d'évaluer mon travail de thèse (j'espère que cela ne sera pas déplaisant). Ainsi que Sidaction, pour avoir financé ma thèse et le projet.

Bien évidemment je remercie tous les membres de mon laboratoire, et tout particulièrement Daniela, sans qui ce travail n'aurait jamais pu aboutir. Pendant ces presque quatre ans tu as été mon guide, mon chef, mon amie (pas toujours dans cette ordre), et je ne saurais jamais assez te remercier du temps que tu as consacré à m'aider dans les manips, m'écouter me plaindre quand tout partait en vrille et m'épauler pour toutes les réflexions scientifiques. Je remercie également Mattéo, mon directeur de thèse, pour m'avoir fait confiance et donné la chance de conduire ce projet, pour tous les bons conseils que tu as pu me donner et le suivi régulier de mon travail. Flore, le laboratoire aurait été bien plus morose si tu n'avais pas été là, toujours là pour discuter (de chats et de lapins, entre autres) et Safi, ma partenaire de laboratoire et de course, nous avons commencé ensemble et finissons ensemble. Les longues heures passées au laboratoire les week-ends, le soir et les jours fériés auraient été bien tristes sans ton sourire et ta bonne humeur. Pierre, le dernier membre et pas des moindres, des intégristes. Toute ma reconnaissance va envers toi pour toutes ces petites choses qui font que tu es toi, ton sens critique, ta bonne humeur, ton dynamisme et les coups de gueules contre les cochons de la salle de culture. Romain, toujours présent pour donner un coup de main ou une oreille compatissante.

Les membres non avoué du 337, Alexis et Camille, pour toutes les petites pauses café et les bons moments à rigoler et se détendre, je suis contente de vous compter parmi mes amis (on se voit au pot de thèse pour trinquer). Je remercie également tous les membres de l'UPR 9002, qui ont contribué à la bonne ambiance générale.

Mon essentiel, Jérôme, toujours présent, quand j'étais triste, fatiguée ou énervée et je n'imagine pas mes journées sans toi. Je ne te remercierai jamais assez pour avoir supporté mes crises de nerfs, m'avoir encouragé quand j'avais des passages à vide et s'être réjoui avec moi, même quand tu ne comprenais pas pourquoi j'étais tellement heureuse d'avoir enfin réussi mon clonage. Pour finir, je remercie toute ma petite famille, (Vanille et Riquita compris), pour tout simplement faire partie de ma vie.

Liste des abréviations

-sssDNA (minus strand strong-stop DNA)
+sssDNA (plus strand strong-stop DNA)
2LTRc (cercle à 2 LTRs)
3TC (lamiduvine)
ALLINI (inhibiteurs allostériques de l'IN)
APOBEC3 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3)
ARNPII (ARN polymérase II)
ARNt$^{Lys\ 3}$ (isoforme 3 de l'ARN de transfert de la lysine)
ATCC (American Type Culture Collection)
AZT (zidovudine)
BAF (Barrier to Autointegration Factor)
CA (capside)
CCD (Catalytic core domain)
CDC (Center for Disease Control and Prevention)
CDK9 (cycline dépendant kinase 9)
CMHI (complexe majeur d'histocompatibilité de classe I)
CPSF6 (cleavage and polyadenylation specificity factor subunit 6)
CRFs (circulating recombinant forms)
CRM-1 (Chromosomal Region Maintenance 1)
CryoEM (cryo-électromicroscopie)
CTD (C terminal domain)
CYCT1 (cycline T1)
CypA (cyclophiline A)
dNTPs (désoxynucléotides)
DTG (Dolutegravir)
Env (enveloppe)
ERAD (endoplasmic réticulum associated dégradation)
ESCRT (endosomal sorting complexes recquired for transport)
EVG (Elvitegravir)
FACT (facilitating chromatine transcription)
FDA (food and drug administration)
Gag (group specific antigen)
Gp41 (glycoprotéine transmembranaire 41)
Gp160 (glycoprotéine de surface 160)
HAART (highly active antiretroviral treatments)
HDGF (Hepatoma Derived Growth Factor)
HEK (human embryonic kidney)
HMGA1 (High Mobility Group Chromosomal Protein A1)
HTLV (Human T-cell Leukemia Virus)
HRP (peroxydase de raifort humaine)
HSPT16 (homologue humain du suppresseur de Ty16)
IBD (domaine de liaison à l'intégrase)
IFNα (interférons α)
IN (intégrase)
INBI (inhibiteurs de la liaison de l'intégrase à l'ADN)
INI1 (Integrase Interactor 1)
INSTI (inhibiteurs de transfert de brin de l'intégrase)
LANL (Los Alamos National Laboratory)
LAV (Lymphadenopathy Associated Virus)
LEDGF/p75 (Lens Epithélium Growth Factor isoform 75)
LEDGIN (inhibiteurs de l'interaction IN-LEDGF/p75)
LT CD4+ (lymphocytes T CD4+)
LTR (long terminal repeat)
MA (matrice)
MMTV (virus de la tumeur mammaire de la souris)
MVV (virus visna-maëdi)
NC (nucléocapside)
Nef (Negative Factor)
NES (signal d'export nucléaire)
NHEJ (non homologous end joining)
NLS (signal de localisation nucléaire)
NNRTI (inhibiteur non nucléosidique de la RT)
NRTI (inhibiteur nucléosidique de la RT)
NTD (N terminal domain)
Nup (Nucléoporine)
PBS (Primer Binding Site)
PCR (Plymerase Chain Reaction)
PI (inhibiteur de la protéase)
PI(4,5)P2 (phopshatidylinositol-4,5-bisphophate)
PIC (complexe de pré-intégration)
PPT (PolyPurine Tract)
P-TEFb (positive transcription elongation factor b)
Pol (polymérase)
PR (protéase)
QPCR (quantitative PCR)
RAL (Raltegravir)

RANBP2 (Ran Binding Protein 2)
Rev (Regulator of Expression of Virion proteins)
RNP (complexe ribonucléoprotéique)
RRE (Rev responsive element)
RSV (virus du sarcome de Rous)
RT (transcriptase inverse)
RTC (complexe de rétro-transcription)
RTPs (produits de la transcription inverse)
SAMHD1 (SAM domain and HD domain-containing protein 1)
SERINC (serine incorporator)
SH3 (sarc homology domain)
SIDA (syndrome d'immunodéficience acquise)
SP1 ou 2 (peptide espace 1 ou 2)
SSRP1 (structure specific recognition protein 1)
TAR (Trans Activation Response)
Tat (Trans-Activator of Transcription)
TFV (tenofovir)
TNPO3 (Transportin 3)
TRIM5α (Tripartite motif-containing protein 5 isoform α)
URFs (unique recombinant forms)
UTR (untranslated terminal repeat)
Vif (Viral Infectivity Factor)
VIH (virus de l'immunodéficience humaine)
VIH-1/M (VIH de type 1 groupe M)
VIH-1/O (VIH de type 1 groupe O)
VIS (virus de l'immunodéficience simienne)
Vpr (Viral Protein R)
Vpu (Viral Protein U)

Table des matières

Introduction

I. Introduction générale sur le VIH

II. L'intégrase

III. Objectifs de l'étude

I. Introduction générale sur le VIH

1. Historique et classification

Durant les trois dernières décennies, le SIDA, syndrome d'immunodéficience acquise, s'est imposé comme un problème majeur de santé mondiale, avec plus de 36 millions de personnes vivant avec le VIH en 2014[1].

En 1981, plusieurs cas de syndrome de Kaposi et d'autres maladies opportunistes ont été rapportés aux Etats-Unis par la CDC (Center for Disease Control and Prevention), suggérant une épidémie de cas d'immunodéficience[2,3]. La maladie est caractérisée par une sévère déficience immunitaire, impliquant une baisse du nombre de lymphocytes T CD4+, accompagnée de nombreuses infections opportunistes. Les premières études suggèrent l'existence d'un agent infectieux transmis par contacts sexuels ou sanguins, suspecté d'origine rétrovirale due aux similitudes (voies de transmission, tropisme de lymphocytes CD4, transcriptase inverse) avec le HTLV (Human T-cell Leukemia Virus) identifié quelques années auparavant[4,5], il est alors appelé HTLV-III.

En 1983, un nouveau rétrovirus, identifié comme l'agent causal du SIDA, a été isolé à partir de ganglions lymphatique d'un patient atteint de multiples lymphadénopathies par l'équipe du Dr. Luc Montagnier de l'institut Pasteur de Paris[6]. Des preuves additionnelles de causalité ont été démontrées en 1984, par une équipe de l'U.S. National Cancer Institute, NIH[7]. Ce nouveau rétrovirus a tout d'abord été appelé LAV (Lymphadenopathy Associated Virus)[8]. En 1986, une souche avec des propriétés biologiques et morphologiques très semblables au LAV mais qui diffère dans certains de ses composants antigénique a été identifiée et nommée LAV-II[9]. Pendant quelques temps, les différents noms (HTLV-III, LAV et LAV-II) ont coexistés, puis la communauté scientifique a adopté les noms de VIH-1 et VIH-2 (virus de l'immunodéficience humaine de type 1 et 2).

Le VIH est un lentivirus qui fait partie de la famille des *Retroviridae*. Ces derniers sont des virus enveloppés et leurs matériels génétiques se composent de deux copies d'ARN monocaténaire de polarité positive. Ils possèdent tous une enzyme avec une activité de transcription inverse (transcriptase inverse, RT) qui permet la synthèse d'un ADN viral à partir du génome ARN.

La famille des Rétrovirus se subdivise en deux sous-familles, les *Orthoretrovirinae* et les *Spumaretrovirinae*. Les *Spumaretrovirinae* sont composés d'un seul genre les spumavirus contenant par exemple le virus de foamy humain non pathogène. La sous-famille des *Orthoretrovirinae* est constituée de 6 genres[10-12] :

- les *Alpharétrovirus* (exemple : virus du sarcome de Rous)
- les *Betarétrovirus* (exemple : virus de la tumeur mammaire de la souris)
- les *Gammarétrovirus* (exemple : virus de la leucémie murine)
- les *Deltarétrovirus* (exemple : virus T-lymphotropique humain)
- les *Epsilonrétrovirus* (exemple : virus de l'hyperplasie épidermique de type 1 du doré jaune)
- les *Lentivirus* (exemple : virus visna-maëdi)

Les *Lentivirus*, auxquels je me suis intéressée au cours de ma thèse, sont caractérisés par des maladies chroniques à période d'incubation longue.

2. Structure du VIH

a. Structure de la particule

Le VIH forme une particule sphérique de 80 à 120 nm de diamètre. L'enveloppe virale est composée de la bicouche lipidique et de protéines de membrane issues des cellules productrices de virions et de glycoprotéines virales. Au niveau de la surface interne, ancrée dans la bicouche lipidique, on trouve la matrice, qui forme un réseau assurant l'intégrité de la particule virale et la capside virale, une structure conique composée de la protéine de capside qui renferme l'ARN, lié à la nucléocapside, ainsi que les enzymes nécessaires à la réplication du virion[13] (Figure **1**).

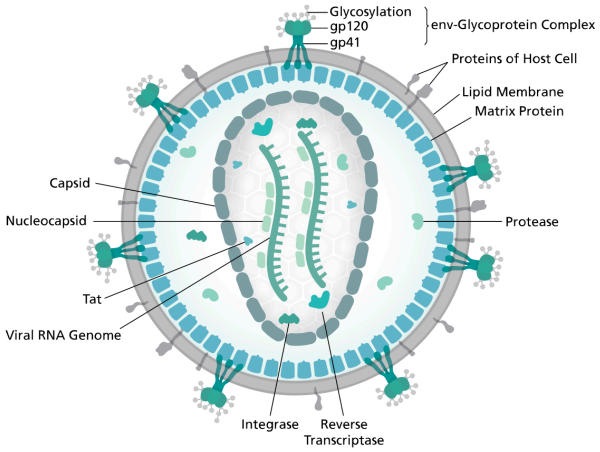


Figure 1 : Structure du VIH, adapté de www.scistyle.com.

b. Organisation du génome viral

La particule virale contient deux copies du génome ARN de polarité positive qui dimérisent, à leur extrémité 5'. Chaque ARN génomique contient les gènes codant pour les protéines d'enveloppes, structurales et enzymatiques communes à tous les rétrovirus, et pour six

protéines additionnelles[14-16] (Figure **2**) :

- les protéines d'enveloppes sont codées par le gène *env* (enveloppe) sous la forme d'un précurseur (gp160), clivé par les enzymes cellulaires[17,18] pour donner la glycoprotéine transmembranaire gp41 et la glycoprotéine de surface gp120.

- les protéines structurales, sont codées par le gène *gag* (group specific antigen) sous la forme du précurseur Pr55Gag qui sera ensuite clivé par la protéase virale pour former la particule mature. Les différentes protéines codées par *gag* sont la matrice (MA), la capside (CA), le peptide espaceur 1 (SP1), la nucléocapside (NC), le peptide espaceur 2 (SP2) et la p6.

- les enzymes virales, sont codées par le gène *pol* (polymérase) sous la forme du précurseur Pr160Gag-Pol (également clivé par la protéase virale) exprimé par un décalage (de - 1) du cadre de lecture à la fin du gène *gag*. Les enzymes sont la protéase (PR), la transcriptase inverse (RT) et l'intégrase (IN).

- les protéines régulatrices Tat et Rev, codées par les gènes du même nom.

- les protéines auxiliaires Vif, Vpr, Vpu (ou Vpx chez VIH-2) et Nef, codées par les gènes du même nom.

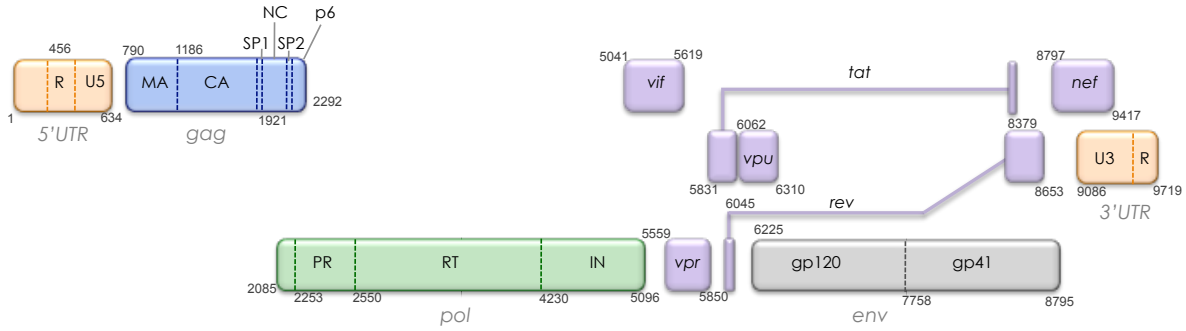


Figure 2 : ARN génomique du VIH. Les trois gènes majeurs (*gag* en bleu, *pol* en vert et *env* en gris) ainsi que les protéines qu'ils codent sont indiqués. Les gènes codant les protéines régulatrices et auxiliaires sont indiqués en violet et les UTR en orange. La numérotation est relative à la souche de référence de laboratoire HXB2. Schéma adapté du HIV compedium[13].

Les extrémités du génome sont composées de régions non traduites (UTR), dupliquées dans la forme provirale intégrée au sein du génome cellulaire, et nommé **LTRs,** pour long terminal repeat[13] (Figure **2**).

Les LTRs sont composés de trois régions :

- R (~100nt), séquence répétée située aux extrémités 5' et 3' de l'ARN génomique, qui joue un rôle important dans la transcription inverse.

- U5 (~80nt), présent dans la région 5'UTR de l'ARN génomique, est essentiel pour la transcription inverse et l'intégration.

- U3 (~450nt), localisé dans la région 3'UTR de l'ARN génomique, contient le promoteur viral, nécessaire pour la réplication[19,20].

c. Les protéines d'enveloppes

L'enveloppe virale est composée de la bicouche lipidique issue de cellules productrices de virions, de protéines de surface cellulaires et d'un complexe de protéines d'enveloppe gp120 et gp41. Comme décrit en amont, les protéines d'enveloppes sont exprimées à partir du précurseur protéique gp160, transcrit à partir d'un ARNm *vpu-env*. Le précurseur est maturé par un clivage endoprotéolytique par la protéase cellulaire furine au sein de l'appareil de Golgi, permettant l'expression à la surface de la particule virale de complexes trimérique de gp41 et gp120[25,26] (Figure **3**). Ceux-ci permettent au virus d'entrer dans les cellules cibles par une liaison au récepteur CD4[21,22], suivi d'une liaison aux corécepteurs CCR5 ou CXCR4[23,24].

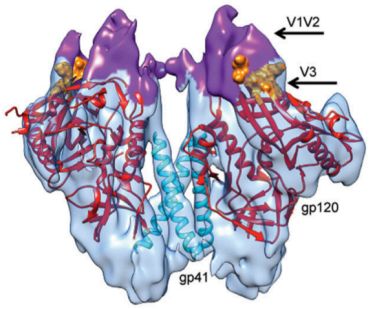


Figure 3 : Structure du complexe protéique d'enveloppes du VIH. La sous-unité gp120 est indiquée en rouge et la sous-unité gp41 en cyan. Les boucles V1V2 et V3 de la gp120 sont illustrées en violet et en orange respectivement. Adapté de [30].

La glycoprotéine de surface **gp120** est une protéine hautement glycosylée de 120kDa[26]. Elle est constituée d'une alternance de régions constantes (C1 à C5) et de régions variables (V1 à V5)[27]. Les régions constantes sont internalisées au sein de la structure alors que les régions variables sont à la surface, reflétant la nécessité pour le virus d'avoir une grande variabilité antigénique dans les parties les plus exposées de la particule virale. Les régions C1 et C5 portent les sites d'interaction avec la gp41 alors que les régions C2, C3 et C4 sont impliquées dans la liaison au récepteur CD4[28]. La liaison de la gp120 avec CD4 entraîne un changement conformationel qui permet le recrutement des corécepteurs CCR5 ou CXCR4[29-31].

La glycoprotéine transmembranaire **gp41** est une protéine de 41kDa, subdivisée en trois domaines, le domaine extracellulaire, la région transmembranaire et la queue cytoplasmique. Le domaine extracellulaire contient un peptide de fusion qui, après un changement conformationel activé par la liaison au récepteur et au corécepteur, est exposé pour permettre la fusion par la formation d'un pont entre les membranes virale et cellulaire[30,32,33].

d. Les protéines structurales

Les protéines structurales, comme décrit précédemment, sont exprimées sous la forme du précurseur Pr55Gag qui sera ensuite clivé par la protéase virale pour libérer les protéines individuelles[34].

La partie externe au cône de capside, sous-jacente à la bicouche lipidique, est appelée matrice et est constituée de trimères de protéine de **matrice** (MA). Cette protéine de 17kDa présente un site de myristylation à son extrémité amino-terminale, nécessaire à l'ancrage stable dans la membrane cellulaire[35-37] et un signal de localisation nucléaire (NLS), important pour l'import nucléaire. De plus, elle interagit avec la queue cytoplasmique de la gp41, assurant l'incorporation de l'enveloppe lors de l'assemblage de la particule virale[38,39].

Le réseau de matrice surplombe la capside virale formée par des multimères de protéines de **capside** (CA). La capside est une protéine de 24kDa composée de deux domaines, un domaine amino-terminal hélicoïdal et un domaine carboxy-terminal, connectés par une région inter-domaine flexible. La protéine de capside s'assemble en hexamères (environ 250) et en 12 pentamères pour former une structure ressemblant à un cône de fullerène (la capside virale) renfermant l'ARN viral lié à la nucléocapside, les enzymes virales et quelques protéines auxiliaires. Lors de l'assemblage en multimères (pentamères ou hexamères) les domaines amino-terminaux se retrouvent exposés à la surface alors que les domaines carboxy-terminal sont internalisés dans la structure (Figure **4**)[40-42].

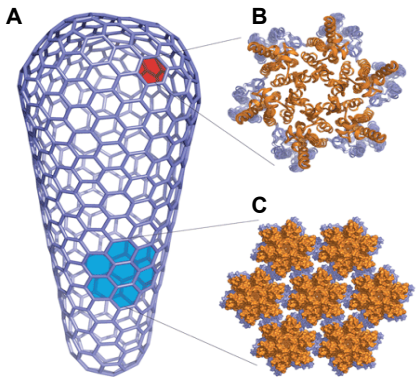


Figure 4 : Schéma du cône de capsides. A. Modèle représentant l'architecture du cône de fullerène formé par les hexamères et pentamères de capsides. **B.** Structure d'un hexamère de capsides avec les domaines N-terminaux à la surface (en violet) et les domaines C-terminaux internalisés (en orange). **C.** Agrandissement de la région surlignée (en bleu) illustrant la structure des hexamères. Adapté de [40].

La protéine de **nucléocapside** (NC) est de petite taille (entre 5 et 7kDa) et comprend deux domaines en doigts de zinc successifs qui lient le Zn2+ avec une haute affinité et permettent la fixation aux acides nucléiques. Cette protéine se lie à l'ARN génomique pour former la nucléocapside, qui est un complexe ribonucléoprotéique (RNP) qui protège le génome viral de la dégradation par les RNases cellulaires, une fois dans le cytoplasme de la cellule. Elle joue également un rôle essentiel au sein du Pr55Gag, en interagissant avec l'ARN génomique (via la séquence spécifique Ψ) afin de permettre son recrutement lors de l'encapsidation[43,44].

La protéine **p6** est un peptide de 6kDa riche en résidus proline et ubiquitinylée après traduction. Elle joue un rôle dans l'association de Pr55Gag à la membrane cellulaire et dans le bourgeonnement des particules néoformées[45-47].

e. Les enzymes virales

Les enzymes virales sont conservées chez tous les rétrovirus, elles sont exprimées sous la forme du précurseur Gag-Pol, qui sera clivé par la protéase.

La **protéase** (PR) est une protéine de 15kDa dont la forme active est un homodimère symétrique. Son site actif est composé des résidus conservés Asp_{25}-Thr_{26}-Gly_{27}, caractéristique de la famille des aspartyl-protéases, dont elle fait partie. Elle est responsable du clivage des précurseurs Gag et Gag-Pol au sein des particules virales néoformées, rendant ainsi les virions infectieux[48-50].

La **transcriptase inverse** (RT) a pour rôle la conversion de l'ARN génomique viral en ADN double brin. Cette protéine est initialement produite sous une forme de 66kDa composée d'un domaine polymérase et d'un domaine RNase H[51,52]. La RT fonctionnelle est un hétérodimère p66/p51, produit du clivage endoprotéolytique du domaine RNaseH sur un homodimère de p66. La sous-unité p66 porte les activités catalytiques, l'activité ADN polymérase, ARN et ADN dépendante et l'activité RNase H qui dégrade spécifiquement l'ARN dans des doubles brins hybrides ARN/ADN. L'analyse cristallographique de la RT montre que le domaine polymérase de la p66 prend une forme ressemblant à une main droite, dont les sous domaines doigts, paume et pouce, qui en font partie, forment un sillon dans lequel se loge le complexe matrice/amorce. Le site actif de l'activité polymérase se situe dans le sous-domaine paume, où les trois résidus Asp_{110}-Asp_{185}-Asp_{186} sont adjacents.

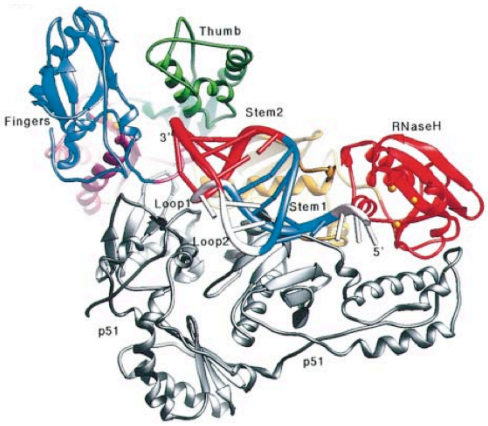


Figure 5 : Structure schématique de la transcriptase inverse en complexe avec l'ADN. La sous-unité p51 de l'hétérodimère est représentée en gris. Les domaines visibles de la sous-unité p66 sont indiqués en couleur, le domaine doigt en bleu, paume en vert et RNaseH en rouge. La double hélice d'ADN est représentée en couleur (bleu-blanc-rouge). Adapté de [54].

Un autre sous-domaine, connexion, relie le domaine RNase H au domaine polymérase. La p51 ne prend pas la même conformation quaternaire que la p66, malgré l'identité de séquence, et n'a pas d'activité catalytique mais exerce un rôle structural important au sein de l'hétérodimère (Figure **5**)[53,54].

L'**intégrase** (IN) est une protéine de 32kDa qui a pour rôle l'intégration de l'ADN viral au sein

du génome de la cellule hôte. Elle est composée de trois domaines distincts, amino-terminal (NTD), cœur catalytique (CCD) et carboxy-terminal (CTD). La forme active de cette protéine est un dimère de dimères asymétriques au sein du complexe de préintégration (PIC) formé par plusieurs protéines virales et cellulaires (MA, CA, RT, IN, NC, ADN viral, LEDGFp75)[55-58]. Cette protéine sera décrite plus en détails dans le **chapitre II**.

f. Les protéines régulatrices

Les protéines régulatrices, Tat et Rev, sont encodées par deux exons et permettent la régulation du cycle viral (Figure **2**).

La protéine **Tat** (Trans-Activator of Transcription) est une protéine de 86 acides aminés, composée de deux domaines, le domaine d'activation, riche en cystéines et le domaine de liaison à l'ARN, riche en arginines. Cette protéine a pour rôle principal l'activation de la transcription des ARN messagers viraux. En effet, au début de la transcription, Tat se lie à l'ARN messager viral naissant au niveau de l'élément TAR (Tat responsive element) situé dans la région 5'UTR de l'ARN[59,60]. Elle recrute le facteur de transcription P-TEFb (Positive Transcription Elongation Factor b) au niveau de l'ARN messager, permettant la phosphorylation de la queue C-terminale de l'ARN polymérase II afin d'activer la transition entre la phase d'initiation et d'élongation de la transcription[61,62].

La protéine **Rev** (Regulator of Expression of Virion proteins) est une protéine de 19kDa. Cette protéine a la particularité de posséder à la fois un signal de localisation nucléaire (NLS) et un signal d'export nucléaire (NES), qui facilite son transfert nucléo-cytoplasmique. Cette protéine se lie à l'ARN messager viral par reconnaissance de la séquence spécifique RRE (Rev responsive element), une séquence ARN structurée de 234 ribonucléotides située au niveau du gène *env*[63,64]. Rev interfère avec le système d'épissage cellulaire en se fixant aux ARN viraux n'ayant pas subi un épissage complet afin de promouvoir leur export vers le cytoplasme grâce à la reconnaissance du récepteur nucléaire CRM-1 (Chromosomal Region Maintenance 1)[65].

g. Les protéines auxiliaires

Le génome du VIH code pour quatre protéines auxiliaires : Vif, Vpr, Vpu et Nef.

La protéine **Vif** (Viral Infectivity Factor) est une protéine de 23kDa qui a pour rôle la suppression de la réponse antivirale médiée par les enzymes APOBEC3 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3)[66,67]. Les enzymes APOBEC3 font partie de la famille des deoxycytidine déaminases, elles suppriment la réplication virale par induction de mutations. Ce facteur de restriction cellulaire sera décrit plus en détails dans la section **I.3.c**. Vif interagit avec APOBEC3 et plusieurs facteurs cellulaires (CBFβ, Elongine

C) pour former un complexe d'ubiquitinylation avec la E2 ubiquitine ligase, afin d'induire la polyubiquitinylation de APOBEC3 et sa dégradation par le protéasome[68].

La protéine **Vpr** (Viral Protein R) est une protéine de 14kDa composée d'un domaine amino-terminal hélicoïdal, riche en résidus acides, d'une région médiane prenant la forme d'un feuillet-β, et d'une région carboxy-terminale, riche en résidus cystéines[69]. Cette protéine joue un rôle dans l'import nucléaire du complexe de préintégration en interagissant avec les karyophérines α (récepteurs cellulaires des signaux NLS) afin d'augmenter leur affinité pour les NLS porté par les protéines du PIC (comme la matrice par exemple)[70,]. Vpr joue également un rôle important de part son habilité dans l'induction de l'arrêt en phase G2 du cycle cellulaire des cellules infectées. Le mécanisme moléculaire d'action de Vpr sur la régulation du cycle cellulaire implique les kinases cellulaires et notamment l'activation de la kinase ATR. L'arrêt du cycle cellulaire en phase G2 permet d'intensifier la réplication du VIH car la transcription virale est plus active pendant cette phase[71].

La protéine **Vpu** (Viral Protein U) est une protéine de 16kDa composée d'un domaine amino-terminal transmembranaire et d'un domaine carboxy-terminal cytoplasmique phosphorylé. Vpu permet le relargage des virions néoformés en inhibant l'expression du facteur restriction cellulaire Tetherine[72,73] (facteur de restriction cellulaire inhibant le relargage des particules virales, voir section **I.3.c).** Elle joue également un rôle dans la dégradation des molécules de CD4 au sein de la cellule, afin d'empêcher la séquestration de la protéine d'enveloppe virale. En effet, les protéines Env et les molécules de CD4 vont transiter par les mêmes compartiments cellulaires, après leur synthèse dans le réticulum endoplasmique, favorisant la séquestration d'Env par CD4 en formant des complexes à l'intérieur de la cellule[74].

La protéine **Nef** (Negative Factor) est une protéine myristylée, localisée dans le cytoplasme et associée à la membrane cellulaire, avec une masse comprise entre 27 et 33kDa. Elle ne possède pas d'activité enzymatique mais exerce néanmoins de nombreuses fonctions au sein de la cellule. Nef module l'expression des molécules à la surface des cellules, telles que CD4, CMH-I (complexe majeur d'histocompatibilité I) et SERINC (serine incorporator), afin de limiter l'effet toxique de l'infection multiple d'une même cellule et de favoriser l'échappement au système immunitaire. Nef est dispensable *in vitro* mais nécessaire *in vivo*, pour atteindre une charge virale élevée. En effet, elle a un effet positif sur la réplication virale et augmente l'infectivité du virus[75-78].

3. Le cycle de réplication du VIH

Le cycle de réplication du VIH débute par la phase précoce avec la reconnaissance du récepteur CD4 et du corécepteur par l'enveloppe, puis par la fusion des membranes virales et cellulaires permettant l'entrée de la capside virale dans la cellule. La décapsidation permet la formation du complexe de rétro-transcription où aura lieu la conversion de l'ARN génomique viral en ADN double brin par la transcriptase inverse, ce complexe s'appellera alors complexe de pré-intégration. Celui-ci sera importé dans le noyau pour permettre l'intégration de l'ADN viral au sein du génome de la cellule hôte. Une fois l'ADN intégré, l'ADN viral reste associé avec le génome cellulaire de façon permanente, c'est pourquoi les infections par le VIH sont à vie.

A ce stade, la phase tardive du cycle débute et l'ARN et les protéines virales sont produits par la machinerie de synthèse cellulaire. Une fois dans le cytoplasme, les précurseurs polyprotéiques Gag et Gag-Pol s'ancrent à la membrane pour assembler une nouvelle particule virale qui bourgeonnera à la surface de la cellule. Enfin, une étape de maturation des précurseurs Gag et Gag-Pol par la protéase virale est nécessaire pour que le virion néoformé soit infectieux (Figure **6**).

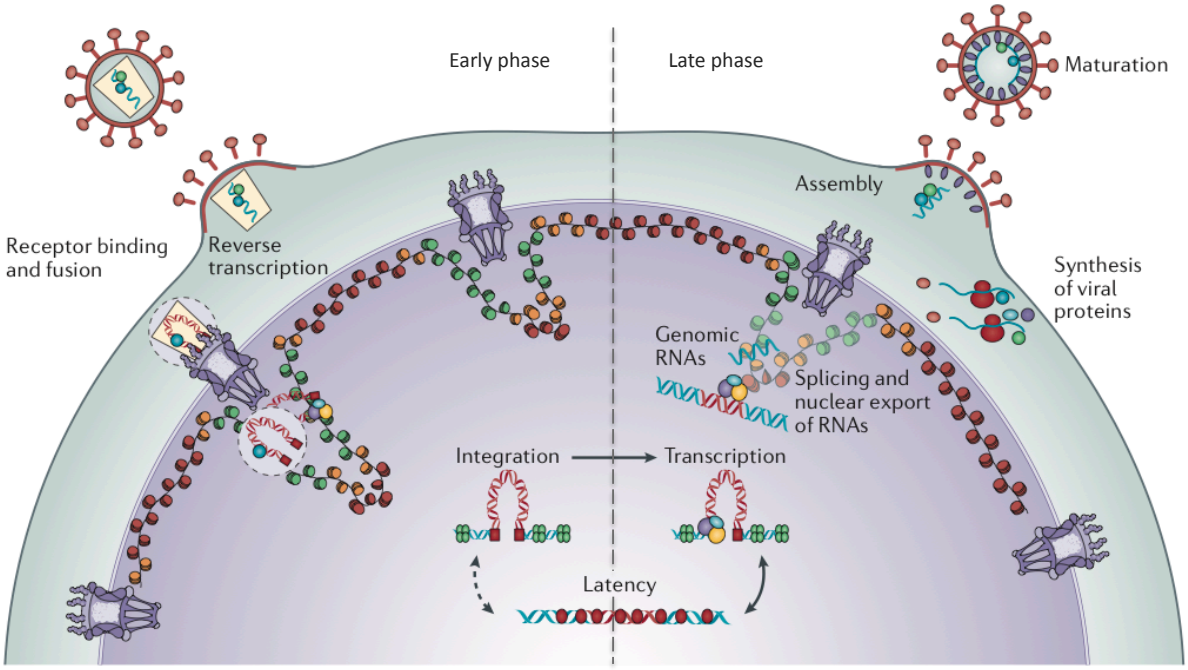


Figure 6 : Cycle de réplication virale. Schématisation du cycle de réplication du VIH avec, à gauche la phase précoce et à droite, la phase tardive. Adapté de [110].

a. Phase précoce

L'**entrée virale** commence par l'adhésion du virus à la membrane cellulaire. Les glycoprotéines d'enveloppes (trimères de gp120 et gp41) s'assemblent pour former des spicules à la surface de la particule virale[30]. L'attachement du virus aux cellules est assuré par plusieurs facteurs. Il peut être aspécifique par l'interaction des spicules d'enveloppe avec les protéoglycanes à la surface de la cellule[79], ou plus spécifique par l'interaction de l'enveloppe avec, soit les intégrines $\alpha 4\beta 7$[80], soit les récepteurs des cellules dendritiques de type DC-SIGN[81]. Dans tous les cas, l'attachement aboutit au rapprochement de l'enveloppe et de son récepteur primaire CD4[82]. La sous-unité gp120 de l'enveloppe se lie à CD4 ce qui provoque un réarrangement de la structure de la gp120, dont un facteur marquant est le repositionnement de la boucle V3, afin de permettre la formation du site de liaison du corécepteur[28].

L'enveloppe du VIH présente deux corécepteurs (CCR5 et CXCR4). Le tropisme est définit selon l'affinité du site de liaison au corécepteur, on parle de tropisme R5 (affinité accrue pour CCR5), et de tropisme X4 (affinité accrue pour CXCR4)[83]. Les souches au tropisme R5 infectent majoritairement les monocytes et les macrophages alors que les souches au tropisme X4 infectent les lymphocytes T CD4+ et induisent la formation de syncytium. Il est admis que lors de la primo-infection et la phase asymptomatique (voir **I.4.a** Physiopathologie de l'infection), les patients sont principalement infectés par des souches au tropisme R5 alors que lors de la phase SIDA, il y a une majorité de souches au tropisme X4[84,85].

Le site de liaison au corécepteur permet la liaison du corécepteur, liaison qui induit un nouveau réarrangement architectural qui libère la sous-unité gp41 et aboutit à l'insertion du peptide de fusion de la gp41 dans la membrane cellulaire[32], amorçant l'étape de fusion par formation d'un faisceau à six hélices[86]. Ce faisceau apporte la force nécessaire à l'enveloppe pour provoquer le rapprochement des deux membranes (virale et cellulaire) et former un pore de fusion par lequel la capside virale va rentrer dans la cellule (Figure **7**)[87].

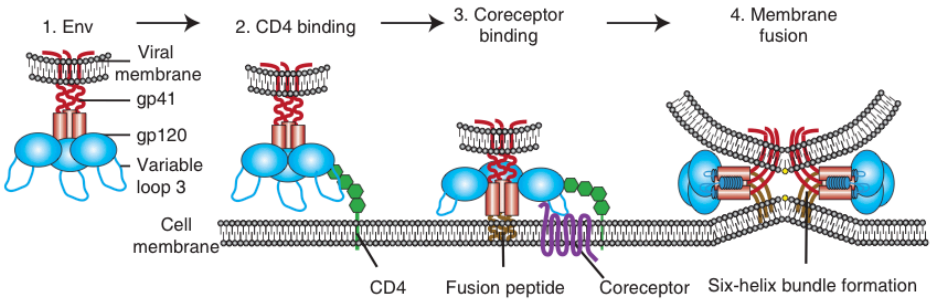


Figure 7 : Aperçu de l'entrée du VIH. 1. L'enveloppe du VIH est composée de trimères de gp120, en bleu et de gp41, en rouge. **2.** Attachement à la cellule par liaison de la gp120 au récepteur CD4. **3.** Liaison du corécepteur grâce à des changements conformationels. **4.** Fusion membranaire suite à l'insert du peptide de fusion de la gp41 dans la membrane cible, suivi de la formation du faisceau à six hélices. Adapté de [87].

Après l'entrée dans le cytoplasme, l'ARN viral est converti en ADN au sein du RTC dans le

but d'être intégré au génome de la cellule. La transcription inverse et l'import nucléaire du PIC[88] sont liés à la formation du RTC, qui découle d'un changement conformationnel de la capside virale appelé étape de **décapsidation**. Il a été démontré que cette étape joue un rôle critique dans le cycle infectieux et permettait de protéger le matériel génétique viral de la dégradation par les protéines cellulaires[89]. Cependant, malgré son importance, la décapsidation est encore peu décrite et son déroulement précis inconnu.

L'analyse du complexe de rétro-transcription, qui se forme dans le cytoplasme, montre qu'il contient plusieurs protéines virales en plus du génome ARN (MA, RT, IN, Vpr), mais présente peu de protéines de capside[90], suggérant que la décapsidation aurait lieu avant ou pendant l'étape de transcription inverse. Une étude récente a proposé un modèle selon lequel la décapsidation est dépendante de la transcription inverse et ne débute que lorsque le premier transfert de brin a eu lieu (voir **I.3.a** transcription inverse)[91].

Malgré la méconnaissance du processus de décapsidation, un facteur cellulaire en interaction avec la capside et jouant un rôle dans cette étape a été identifié, la cyclophiline A (CypA). La CypA est une peptidyl-isomérase qui est internalisée dans les particules lors de l'assemblage grâce à son interaction avec le domaine CA du précurseur Gag[92]. Elle prévient, dans certains cas, la décapsidation prématurée[93] et facilite l'engagement des facteurs cellulaires tels que TNPO3 (Transportin 3), CPSF6 (cleavage and polyadenylation specificity factor subunit 6) et des nucléoporines (Nup153, Nup358) qui interviennent dans l'étape de décapsidation et l'entrée dans le noyau[89,94].

La **transcription inverse** consiste en une série complexe de réactions biochimiques qui se termine avec la génération de l'ADN linéaire double brin à partir de l'ARN génomique viral simple brin, qui sera intégré dans le génome de l'hôte. Cette étape du cycle est catalysée par la RT au sein du RTC et a lieu dans le cytoplasme de la cellule infectée. La RT utilise l'ARN génomique comme matrice et l'isoforme 3 de l'ARN de transfert de la lysine ($ARNt^{Lys\ 3}$) comme amorce[95]. Plusieurs études ont montré que cet ARN de transfert (ainsi que deux autres isoformes, la 1 et la 2) étaient présent dans la capside virale des particules néoformées, grâce au recrutement par le domaine NC du précurseur Gag[96,97]. Seul l'isoforme 3 de l'$ARNt^{Lys}$ peut servir d'amorce pour la RT, les deux autres isoformes joueraient un rôle en aval, lors de l'import nucléaire[98]. La synthèse débute par l'hybridation de l'$ARNt^{Lys\ 3}$ sur sa séquence complémentaire, le PBS (Primer Binding Site), de 18 nucléotides situé à l'extrémité 5' de l'ARN génomique viral (Figure **8.A**). Cette liaison est un mécanisme complexe qui fait intervenir la NC et son activité chaperonne, qui va favoriser le rapprochement de l'amorce sur le site d'hybridation et déstabiliser la structure de l'ARN génomique afin de permettre l'hybridation[98-100].

L'amorce est allongée de 3' vers 5' jusqu'à atteindre l'extrémité 5' de l'ARN, afin de générer un court brin d'ADN négatif appelé minus strand strong-stop DNA (-sssDNA), possédant les séquences complémentaires de R et U5 (Figure **8.B**). L'ARN complémentaire du -sssDNA se retrouve hybridé à de l'ADN, qui est une matrice idéale pour la RNase H. Le -sssDNA est libéré grâce au clivage de l'ARN par la RNaseH, permettant son transfert sur l'extrémité 3' de l'ARN génomique grâce à sa complémentarité avec la séquence répétée R.

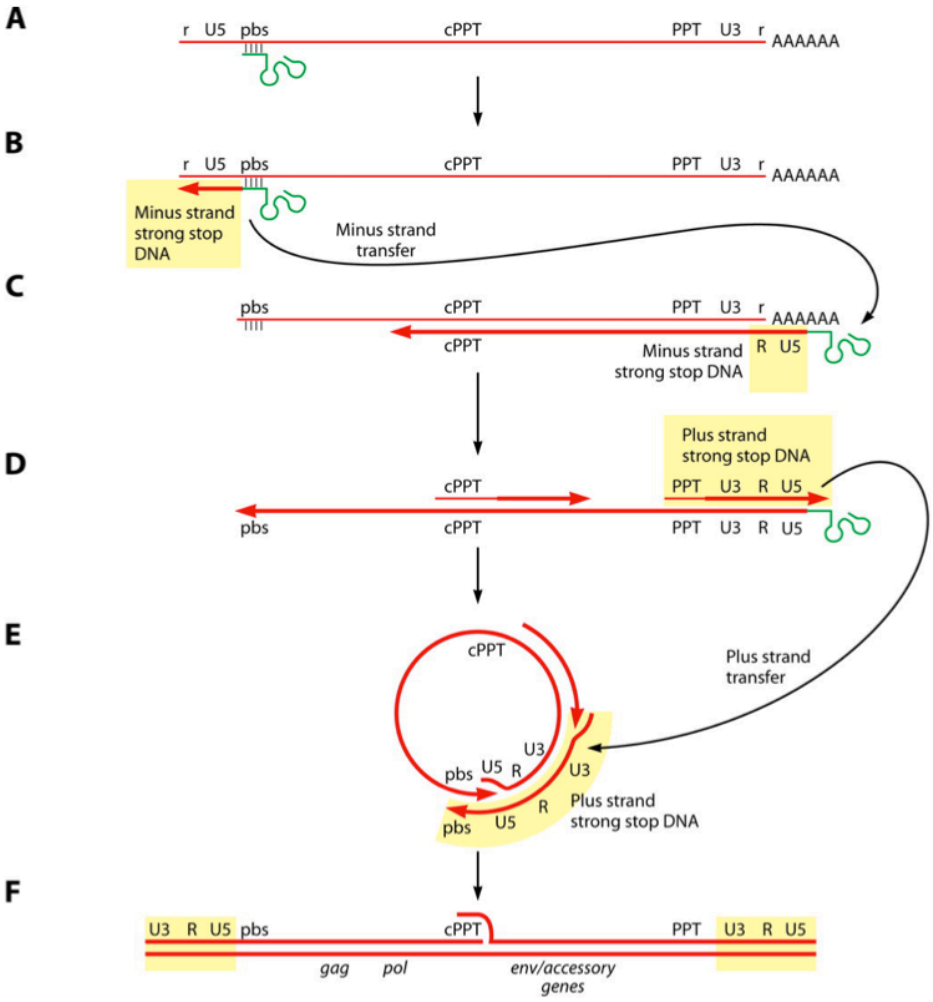


Figure 8 : Mécanisme de la transcription inverse. A. L'amorce ARNt^{Lys,3} (en vert) s'hybride à sa séquence complémentaire sur l'ARN génomique viral (fine ligne rouge), le site de liaison de l'amorce (PBS). **B.** La synthèse du brin négatif d'ADN (épaisse ligne rouge) commence par l'ARNt et continue jusqu'à atteindre l'extrémité 5', générant le minus strand strong-stop DNA (-sssDNA). **C.** Après la dégradation du brin d'ARN hybridé à l'ADN par la RNaseH, le -sssDNA est libéré et peut s'hybrider sur la séquence complémentaire R à l'extrémité 3'. La synthèse du brin négatif d'ADN continue et l'ARN hybridé à l'ADN est dégradé en parallèle. **D.** Au fur et à mesure de la synthèse du brin négatif, la synthèse de l'ADN du brin positif est initiée à partir des séquences PPT, résistantes à la RNaseH. La synthèse à partir du PPT en 3' donne naissance au plus strand strong-stop DNA (+sssDNA). **E.** Après le clivage de l' ARNt^{Lys,3} hybridé au PBS par la RNase H, le +sssDNA s'hybride sur la région PBS complémentaire du brin négatif, afin de permettre la synthèse complète des LTRs. **F.** La synthèse aboutit à un ADN bicaténaire avec un LTR à chaque extrémité, séparé par un court chevauchement appelé le flap central au niveau de la région PPT central (cPPT). Adapté de [182].

Comme le VIH contient deux copies d'ARN, le transfert de brin peut être intermoléculaire ou intramoléculaire. Lorsque les deux copies d'ARN viral ne sont pas identiques, le transfert de brin intermoléculaire génère une variabilité par recombinaison (voir **I.5.b**)[101].

La synthèse du brin négatif se poursuit par l'élongation du -sssDNA et l'ARN complémentaire est dégradé par l'activité de la RNase H, au fur et à mesure de la synthèse (Figure **8.C**). Bien que presque tout le génome ARN soit clivé par la RNase H, deux régions riches en purines appelées PPT (PolyPurine Tract) résistent à la dégradation par la RNase H. L'une se situe au centre de la région codante de l'IN, le PPT central, et l'autre dans la 3' UTR (extrémité 5' d'U3), le 3'PPT. Les régions PPT vont servir d'amorce pour la synthèse du brin positif de l'ADN viral.
Pendant la synthèse du brin négatif, la synthèse du brin positif est donc initiée à partir des régions PPT. Le fragment d'ADN positif résultant de la synthèse du brin complémentaire de l'ARNt$^{Lys\ 3}$ est appelé plus strand strong-stop DNA (+sssDNA) (Figure **8.D**). Après le clivage de l'ARNt hybridé au PBS par la RNase H, le +sssDNA s'hybride sur la région PBS complémentaire du brin négatif, afin de permettre la synthèse complète des LTRs (Figure **8.E**). La synthèse aboutit à un ADN double brin constitué de deux segments distincts, séparés par un court chevauchement appelé le flap central au niveau de la région PPT central (Figure **8.F**)[94,98-100].

Durant la progression de la transcription inverse, le RTC se déplace grâce aux microtubules[102] vers les pores nucléaires, afin de permettre l'**import nucléaire**. Il est admis que l'entrée du complexe de pré-intégration (PIC) dans le noyau est un processus actif gouverné par plusieurs facteurs viraux et cellulaires. La protéine de capside est un déterminant majeur de l'entrée, ce qui supporte l'hypothèse que la décapsidation et la transcription inverse se produisent aux alentours des pores nucléaires[103]. Le PIC est un dérivé du complexe de rétro-transcription, contenant l'ADN viral double brin, les protéines virales du RTC et plusieurs facteurs cellulaires. La cyclophiline A lie une boucle exposée de la protéine de capside, et favorise le recrutement de RANBP2 (Ran Binding Protein 2), appelée aussi Nup358, qui permet l'attachement du PIC au pore nucléaire. RANBP2 forme avec la nucléoporine 153 (Nup153) un site d'attachement aux molécules qui traversent le pore nucléaire (Figure **9**)[108,109].
D'autres facteurs cellulaires tels que TNPO3 et CPSF6 sont recrutés pour promouvoir l'import nucléaire[104,105]. Comme décrit précédemment, le PIC contient plusieurs protéines virales qui portent un signal de localisation nucléaire (MA, LEDGF) et sont impliquées dans l'interaction avec les protéines du pore nucléaire[106,107].

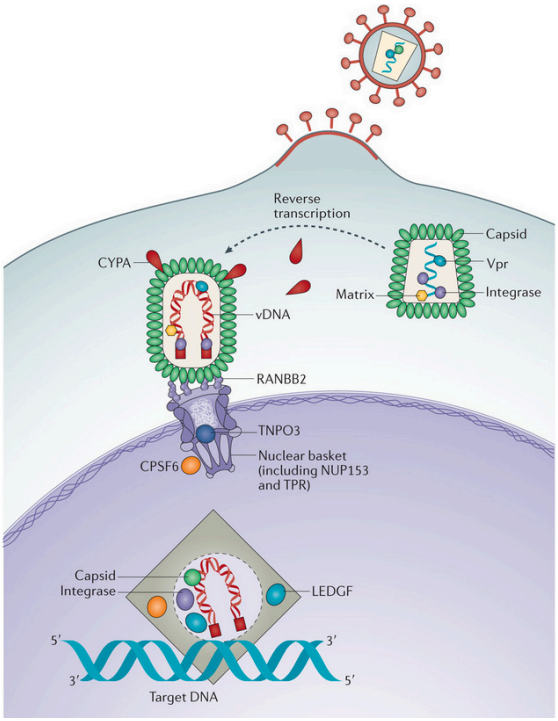


Figure 9 : Import nucléaire. Schématisation des étapes précédant l'intégration : transcription inverse, import nucléaire et ciblage de la chromatine. Adapté de [110].

Une fois dans le noyau, l'ADN viral est **intégré** au sein du génome de la cellule hôte, et est alors nommé provirus. Ce procédé est lié à l'import nucléaire car plusieurs protéines qui interviennent dans l'entrée (Nup153, CPSF6) gouvernent, avec le cofacteur cellulaire LEDGF/p75 (Lens Epithélium Growth Factor isoform 75), la sélectivité du site d'intégration[110]. Cette étape du cycle sera décrite plus en détails dans la partie **II.2** mécanisme d'action.

b. Phase tardive

L'intégration de l'ADN au sein du génome cellulaire marque la transition entre la phase précoce et la phase tardive du cycle de réplication (Figure **6**). A ce stade, le provirus se comporte comme un gène cellulaire et sera transcrit en ARN messager à partir du 5' LTR, qui agit comme promoteur, jusqu'au 3'LTR qui contient le signal de polyadénylation, terminateur de la transcription.

La **transcription** est initiée en aval de la séquence U3, qui contient le promoteur viral et plusieurs autres séquences régulatrices de la transcription : des sites de fixation de facteurs de transcription cellulaires, dont trois pour le facteur de transcription cellulaire Sp1, qui permettent le recrutement de l'ARN polymérase II (ARNPII) sur le promoteur, d'autres pour les facteurs NF-κB et AP-1 et la séquence TAR (Trans Activation Response)[111,112]. Après liaison de l'ARNPII sur le promoteur, la protéine Tat associée au facteur de transcription P-TEFb (positive transcription elongation factor b) se lie à la séquence TAR sur les ARN messagers viraux néoformés[59-62]. P-TEFb est un complexe composé de la cycline T1 (CYCT1) et de la kinase cycline dépendante 9 (CDK9), qui active l'élongation de la transcription par phosphorylation de la queue carboxy-terminale de l'ARN polymérase II[113]. Il

a été démontré que le facteur NF-κB peut se substituer partiellement à l'action de Tat, cependant l'infection reste moins efficace, démontrant l'importance de ce système spécifique de régulation Tat/TAR[114].

La transcription du provirus aboutit à un ARN messager d'une taille d'environ 9kpb, qui contient de nombreux sites donneurs et accepteurs d'épissages. Ainsi, une série d'épissages alternatifs permet d'aboutir à la formation de plusieurs transcrits, dont la plupart sont partiellement ou non épissés et seraient normalement retenus dans le noyau par la machinerie cellulaire, sans l'action de la protéine Rev[115,116]. En effet, les ARN messagers viraux totalement épissés sont exportés dans le cytoplasme par la voie canonique qui permet l'expression des protéines Tat, Rev et Nef, alors que les transcrits partiellement ou non épissés nécessitent la liaison de Rev pour être exportés. Cette protéine se lie aux ARN messagers viraux par reconnaissance de la séquence RRE, et permet l'export des transcrits par la voie CRM-1[117]. Rev contient un NES (signal d'export nucléaire) qui permet, une fois lié à l'ARN messager, de recruter le complexe CRM-1/RAN-GTP afin d'être exporté dans le cytoplasme[118,119].

La traduction de ces messagers permet l'expression du précurseur d'enveloppe gp160, du précurseur Pr55Gag et, par un décalage de -1 du cadre de lecture qui se produit tous les 20 évènements de traduction, l'expression du précurseur Pr160Gag-Pol, au sein du réticulum endoplasmique rugueux[120]. Le précurseur gp160 est glycosylé au niveau du domaine gp120 et va transiter par l'appareil de Golgi où il sera clivé par la protéase cellulaire furine, afin de former les spicules d'hétérodimères de gp120 et gp41, qui seront ensuite ancrés dans la membrane plasmique grâce à la gp41[86]. En parallèle, la protéine virale Vpu induit la destruction des protéines de surface CD4 piégées dans le réticulum endoplasmique due à leur liaison au domaine gp120 de l'enveloppe virale. En effet, Vpu se lie aux complexes CD4-env afin d'éviter la séquestration de l'enveloppe à l'intérieur de la cellule en activant la dégradation des CD4 dans le protéasome par la voie ubiquitine-dépendante[121,122].

L'**assemblage** des virions est un processus complexe, qui prend place à la périphérie de la membrane plasmique et pour lequel chaque domaine de Gag (MA, CA, NC, p6) joue un rôle crucial[123,124]. L'extrémité N-terminale myristylée de MA est exposée grâce au changement de conformation provoqué par l'interaction de MA avec le phopshatidylinositol-4,5-bisphophate (PI(4,5)P2), permettant ainsi l'ancrage de Gag dans la membrane[125,126]. Les précurseurs Pr55Gag et Pr160Gag-Pol s'accumulent à la membrane et multimérisent pour former une structure immature, stabilisée par des interactions Gag-Gag au niveau du domaine CA[127]. La plupart des monomères Gag qui s'ancrent à la membrane sont liés à l'ARN grâce à la forte affinité du domaine NC pour les acides nucléiques. L'ARN génomique viral est

spécifiquement reconnu par NC grâce à la séquence d'encapsidation Ψ, située dans la 5'UTR[44,123,128]. Les spicules de glycoprotéines d'enveloppe atteignent la membrane plasmique indépendamment de Gag mais sont incorporés aux virions naissant grâce à l'interaction de la gp41 avec le domaine MA de Gag[129].

Bien que Gag soit responsable de l'assemblage des virions, le virus usurpe la voie cellulaire ESCRT (endosomal sorting complexes recquired for transport) pour bourgeonner et libérer les virions immatures, grâce au recrutement des facteurs Tsg101 et ALIX par le domaine p6 de Gag[130-132]. Brièvement, Tsg101 permet de recruter le complexe ESCRT I et fonctionne en amont avec ALIX afin de recruter le complexe ESCRT III pour permettre la fission membranaire, et le complexe Vsp4 pour le recyclage des facteurs ESCRT[133] (Figure **10**).

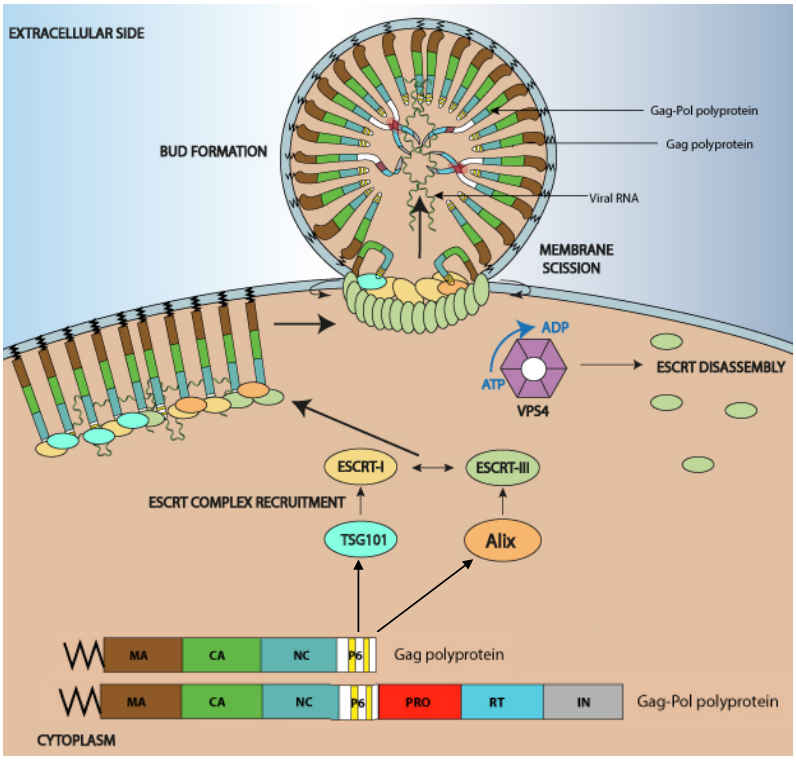


Figure 10 : Assemblage et bourgeonnement de la particule virale. Schématisation du rôle de la p6, en tant que composant des précurseurs Gag et Gag-Pol, dans l'assemblage de la particule via le recrutement du complexe cellulaire ESCRT. Adapté de [134].

Les virions néoformés nécessitent une étape de **maturation** avant d'être infectieux, assurée par la protéase (PR) virale, une protéine de la famille des aspartyl-protéases, comme décrit en amont[135]. Son site actif est un dimère et chaque sous-unité du dimère contribue à la catalyse. La protéase peut s'autocliver au sein de la particule virale grâce à la dimérisation des Pr160Gag-Pol, qui permet de reconstituer le site actif des domaines PR[136,137]. La protéase libère, par dix clivages protéolytiques, les protéines structurales (Gag) et les enzymes matures (Pol), grâce à la reconnaissance de sites spécifiques[138]. La maturation est un processus dynamique en plusieurs étapes qui implique une série de changements de

conformations afin d'aboutir à la structure finale de la particule. La morphogénèse correcte est assurée par un contrôle temporel des clivages[139], qui permet d'assurer la formation du cône de capside entouré du réseau de matrice, contenant la nucléocapside complexée à l'ARN et les protéines virales (Figure **11**).

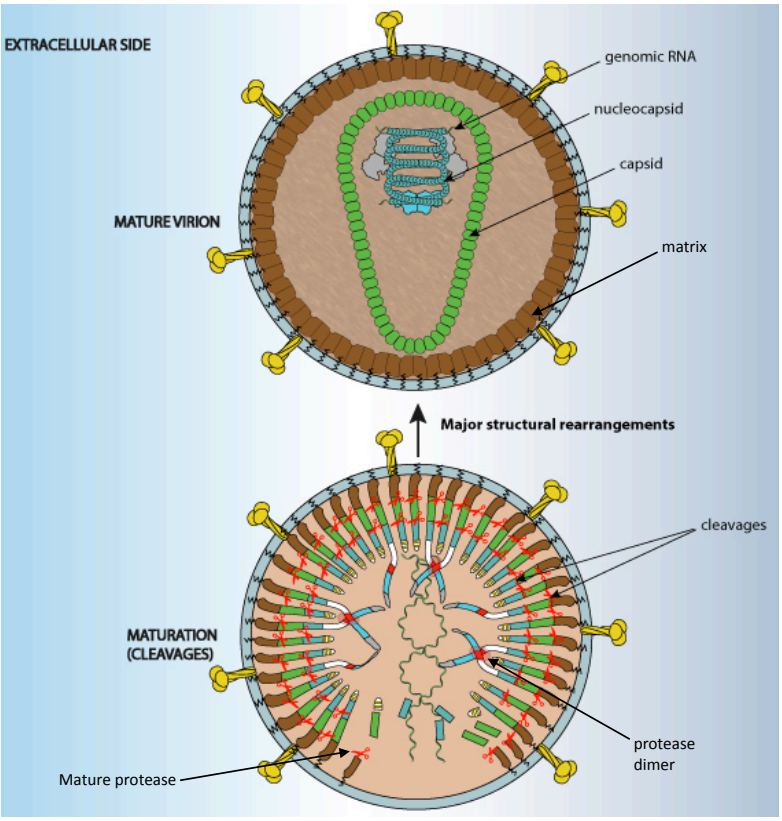


Figure 11 : Maturation des particules virales. Les précurseurs Gag-Pol se dimérisent afin de permettre l'autoclivage de la protéase mature, afin que celle-ci clive les précurseurs Gag et Gag-Pol, libérant les protéines individuelles pour la bonne morphogénèse de la particule. Adapté de [134].

c. Restriction cellulaire et évasion immunitaire

Bien que la cellule procure tous les facteurs nécessaires à la production des virions, plusieurs gènes cellulaires codent pour des protéines antivirales que l'on appelle facteurs de restriction cellulaire, dont l'expression est stimulée par les interférons α (IFNα). Cependant le VIH a développé un grand nombre de stratégies pour éviter et antagoniser la restriction cellulaire.

L'une des premières étapes du cycle viral ciblée par les facteurs de restriction et l'entrée avec les protéines de la famille des **SERINC** (serine incorporator). Ce sont des protéines transmembranaires capables d'incorporer des sérines dans les phosphatidylsérines et les sphingolipides, affectant ainsi la composition lipidique de l'enveloppe virale, essentielle à l'infectivité. En effet, il a été démontré que SERINC3 et SERINC5 affectent la fusion membranaire dans les cellules cibles et donc, le relargage de la capside dans le cytoplasme.

La protéine virale Nef régule l'expression de SERINC3 et SERINC5 à la surface des cellules productrices de virions afin de neutraliser leurs effets[140,141].

Un autre facteur cellulaire, **TRIM5α** (Tripartite motif-containing protein 5 isoform α) provoque une décapsidation prématurée, inhibant ainsi les étapes suivantes de la phase précoce (transcription inverse et import nucléaire) et l'établissement du provirus. TRIM5α forme des trimères qui peuvent contacter les hexamères de CA, et engendrer des lésions irréversibles dans la structure de la capside virale (Figure **12.A**). Le mécanisme d'action de ce facteur est encore peu décrit, mais l'on sait que le VIH contrecarre l'action de TRIM5α grâce au recrutement de la cyclophiline A cellulaire (CypA), qui est en compétition avec TRIM5α pour la liaison à la protéine de capside, et diminue donc son effet restrictif[142,143].

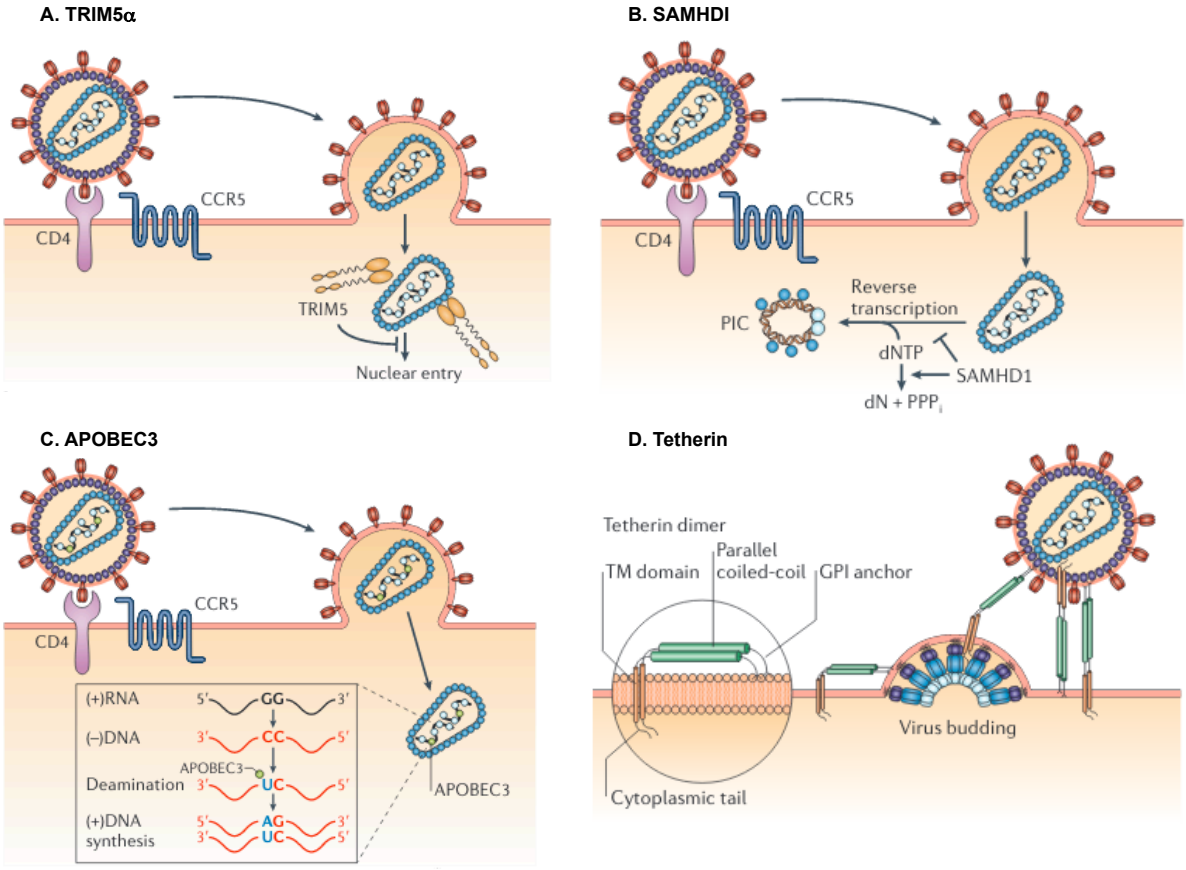


Figure 12 : Facteurs de restriction cellulaire. A. Représentation schématique de TRIM5α et de son action de restriction. **B.** Restriction par SAMHD1 lors de la transcription inverse. **C.** Substitution des C en U par APOBEC3 à l'intérieur de la capside. **D.** Rétention des particules bourgeonnantes à la surface de la cellule par Tetherine. Adapté de [154].

Les cellules myéloïdes sont des cibles naturelles des lentivirus, VIH compris. Cependant, certaines cellules myéloïdes (cellules dendritiques, monocytes dérivés de macrophages) sont résistantes à l'infection par le VIH, due à l'expression du facteur de restriction **SAMHD1**

(SAM domain and HD domain-containing protein 1). Cette protéine est une dNTPase dépendante du GTP et permet la régulation du métabolisme des désoxynucléotides (dNTPs). Elle inhibe la transcription inverse en maintenant le taux de dNTPs dans la cellule tellement bas que la transcription inverse est défavorisée (Figure **12.B**). C'est la raison pour laquelle cette restriction n'est exprimée que dans un type cellulaire précis, car elle affecte également la synthèse cellulaire. Il n'y a pas de défense virale chez le VIH de type 1, bien que la protéine auxiliaire Vpx exprimée chez le VIH de type 2 et plusieurs virus de l'immunodéficience simienne (VIS) (voir **I.5.a** Phylogénie), induit la dégradation de SAMHD1.

Une autre enzyme bloque le VIH au niveau de la transcription inverse, **APOBEC3**. Cette enzyme cellulaire fait partie de la famille des cytidines déaminases, composée de sept membres (A à H), dont seul APOBEC3G (A3G) et APOBEC3F (A3F) ciblent le VIH. La protéine A3G (ou A3F) est incorporée dans les virions et catalyse la mutation des C en U, engendrant une hypermutation de l'ADN lors de la synthèse du brin négatif[146] (Figure **12.C**). La protéine virale Vif est capable d'empêcher l'incorporation d'A3G/F dans les particules naissantes en le liant dans le cytoplasme. Vif recrute ensuite le complexe E3 ubiquitine ligase grâce à son site de liaison avec un membre de ce complexe, la Culline 5, aboutissant à la dégradation d'A3G/F. En effet ce complexe cellulaire est impliqué dans la régulation des protéines par la voie de dégradation du protéasome[68,147,148].

Un autre facteur de restriction est la **Tetherine** appelée aussi **BST-2,** qui bloque une étape de la phase tardive du cycle. Cette protéine possède deux domaines transmembranaires qui s'ancrent dans la membrane plasmique et dans la membrane virale lors du bourgeonnement viral. Elle agit alors comme une ancre qui bloque les virions à la surface de la cellule et empêche leur relargage[149] (Figure **12.D**). Cette protéine est contrée chez le VIH-1 par Vpu, alors que c'est la protéine d'enveloppe chez VIH-2 qui permet d'antagoniser ce facteur, avec un mécanisme similaire à celui de Vpu[150]. Il a été démontré que Vpu régule la présence de Tetherine à la surface de la cellule en provoquant son ubiquitinylation, qui mène à son endocytose afin d'être dégradé dans les lysosomes.[151].

Vpu joue un autre rôle dans la régulation de l'expression virale en empêchant la formation de complexes env-CD4 à la sortie du réticulum endoplasmique. En effet, d'autres mécanismes cellulaires ont une action antivirale. Le transit des molécules CD4 dans les mêmes compartiments cellulaires que la protéine d'enveloppe provoque, en l'absence de Vpu, la séquestration des spicules d'enveloppe, via l'interaction de CD4 avec la gp120. La dégradation de CD4 enclenchée par Vpu détourne deux voies de dégradation cellulaire, la voie du protéasome ubiquitine dépendante et la voie de dégradation associée au réticulum

endoplasmique (ERAD)[74]. La protéine Nef cible également les récepteurs CD4 à la surface de la cellule afin d'éviter leur interaction avec la gp120. En effet, en absence de Nef, le taux de CD4 à la surface de la cellule est plus élevé et un grand nombre se lie aux spicules d'enveloppes qui bourgeonnent à la membrane, empêchant ainsi le relargage des nouveaux virions[152].

En plus de participer à la régulation du cycle viral, la protéine Nef permet également l'évasion du système immunitaire de l'hôte. En effet, il a été démontré que certains anticorps dirigés contre l'enveloppe reconnaissaient des épitopes exposés uniquement lorsqu'elle est en interaction avec CD4, la régulation de CD4 par Nef favoriserait donc l'échappement à ce mécanisme immunitaire[153]. De plus, de la même façon qu'elle régule les CD4 à la surface de la cellule, Nef cible également les complexes majeurs d'histocompatibilité de classe I. Ces molécules permettent l'exposition des peptides viraux à la surface de cellules infectées, qui sont alors reconnues et dégradées par les lymphocytes T cytotoxiques[154].

4. Physiopathologie et traitement de l'infection

Le VIH excelle dans le détournement des voies cellulaires de l'hôte, tout en échappant aux composants antiviraux du système immunitaire. De ce fait, l'infection se caractérise par un déficit graduel de la population de lymphocytes T CD4+ et son diagnostic est basé sur la détection d'anticorps et d'antigènes spécifiques.

a. Physiopathologie de l'infection

En absence de traitement, l'infection se déroule en trois phases[156-158]: la primo-infection, la phase asymptomatique ou latente et la phase symptomatique ou SIDA.

La **primo-infection,** dont les premiers symptômes sont apparentés à un syndrome pseudo grippal, survient durant les premières semaines post infection. Cette phase se caractérise par une réplication virale intense associée à une baisse de la quantité des lymphocytes T CD4+ (LT CD4+).

La phase de latence ou **asymptomatique** est de durée variable allant de quelques mois à plusieurs années, d'où le terme de latence. La réplication du virus se stabilise à un niveau basal, concomitant avec la stabilisation de la diminution des LT CD4+.

En général, les symptômes caractéristiques apparaissent dans les 10 ans après l'infection. La phase **SIDA** (syndrome d'immunodéficience acquise) est déclarée dès l'augmentation de la virémie, couplée avec une déficience en LT CD4+. Ce déficit immunitaire favorise l'apparition de maladies opportunistes et de tumeurs, aboutissant en général à la mort du patient.

b. Traitements antirétroviraux

Plusieurs inhibiteurs rétroviraux sont actuellement approuvés par la US food and drug administration (FDA), dont les cibles thérapeutiques sont multiples (entrée, transcription inverse, intégration et maturation)[159].

Les inhibiteurs de l'**entrée virale** sont de deux catégories. La première vise à inhiber la liaison au récepteur et corécepteur de la cellule, et à l'heure d'aujourd'hui, seul un inhibiteur nommé maravaroc circule sur le marché pharmaceutique. Maravaroc se lie au corécepteur CCR5 et provoque un changement de conformation qui empêche la liaison de la gp120. L'autre catégorie cible la fusion membranaire effectuée par la gp41. L'inhibiteur enfurvitide, approuvé par la FDA, interagit avec le faisceau à 6 hélices de gp41 afin d'empêcher la fusion des membranes virale et cellulaire[160].

Les inhibiteurs de la **transcription inverse** sont classés en deux groupes, les inhibiteurs nucléosidiques de la RT (NRTI) et les inhibiteurs non nucléosidiques de la RT (NNRTI). Les NRTI sont des analogues nucléosidiques qui nécessitent plusieurs étapes de phosphorylation par la machinerie cellulaire pour être actifs. Sous sa forme triphosphate, le NRTI est en compétition avec les désoxynucléotides cellulaires pour l'incorporation par la RT. Une fois incorporé dans la chaîne peptidique, le NRTI agit comme un terminateur de chaîne, mettant fin à la synthèse de l'ADN viral. Le premier NRTI approuvé par la FDA est la zidovudine (AZT). Plusieurs autres ont ensuite suivi, comme le tenofovir (TFV) ou la lamiduvine (3TC)[161].
Contrairement aux analogues nucléosidiques, les NNRTI se lient de manière spécifique et non compétitive à une région hydrophobique de la RT, proche du site actif de la polymérase. L'interaction du NNRTI avec la RT provoque un réarrangement structural qui altère le fonctionnement de l'enzyme. De part leur caractère spécifique, les NNRTI ne sont pas toxiques pour la cellule car il n'y a aucun risque d'altération de la synthèse d'ADN cellulaire contrairement aux NRTI. Cependant leur utilisation clinique a été limitée due à l'émergence de nombreuses souches résistantes. Quelques exemples de molécules sont la tuvirapine et la névirapine[162].

Les inhibiteurs de l'**intégration** sont classés en plusieurs catégories. Il y a des inhibiteurs allostériques de l'intégrase et des inhibiteurs catalytiques, comme par exemple les inhibiteurs de transfert de brin (INSTI), dont seulement trois ont été approuvés par la FDA, raltegravir (RAL), elvitegravir (EVG) et dolutegravir (DTG). Ces inhibiteurs seront plus largement décrits dans la partie **II.4** Inhibiteurs de l'intégrase[163].

Les inhibiteurs de la **maturation,** ou inhibiteur de la protéase (PI) **,** sont peptidomimétique, ce qui signifie qu'ils imitent le substrat naturel de la protéase virale pour se lier, en compétition avec le substrat naturel, dans le site actif. Plusieurs inhibiteurs de la protéase sont approuvés par la FDA avec notamment saquinavir, qui est le premier identifié, ou encore nelfinavir. Cependant leur utilisation a induit l'émergence de nombreuses mutations de résistances et ils sont, de nos jours, uniquement utilisés en association avec d'autres inhibiteurs dans le cadre de la HAART[164].

L'approche thérapeutique HAART (highly active antiretroviral treatments) consiste en la combinaison de plusieurs antirétroviraux pour inhiber la réplication virale jusqu'à un taux trop faible pour permettre l'émergence de nouveaux variants du virus, résistants aux antirétroviraux. Classiquement elle combine deux inhibiteurs nucléosidiques de la RT avec un inhibiteur non nucléosidique de la RT, additionné d'un inhibiteur de la protéase ou plus récemment d'un inhibiteur de l'intégrase[165].

5. Diversité génétique du VIH

Le VIH dispose d'une grande diversité génétique, qui est un des obstacles majeurs à son éradication. Comme mentionné précédemment, ce virus a la capacité de générer rapidement un grand nombre de variants génétiques qui permet la sélection de nombreuses résistances aux antirétroviraux[166]. Cependant, cette capacité remarquable d'hypermutations n'est pas la seule source de variabilité puisque plusieurs groupes phylogénétiques ont été décrits.

a. Origines phylogénétiques

Le VIH prend pour origine plusieurs **transmissions zoonotiques** différentes des virus du singe à l'Homme, ce qui explique en partie sa diversité génétique. Tout d'abord, il y a deux types de VIH qui ont été identifiés, le VIH de type 1 (VIH-1) et le VIH de type 2 (VIH-2)[167]. Ceux-ci se différencient par leur origine simienne, le VIH-2 prend pour ancêtre le virus de l'immunodéficience simienne (VIS) du mangabey enfumé et est divisé en 8 groupes (A à H), dont le A et le B sont les plus importants[168]. Le VIH-2 ne sera pas plus amplement décrit car cette étude porte sur le VIH-1.

La souche de VIS de la sous-espèce *Pan troglodytes troglodytes* des chimpanzés (VIScpz) a été sujette à des transmissions inter-espèces, donnant naissance au virus de l'immunodéficience simienne du gorille (VISgor) et au VIH-1 chez l'Homme[169]. Le VIH-1 est divisé en quatre groupes phylogénétiques (M, N, O et P) qui reflètent des évènements de transmissions zoonotiques indépendantes. Les groupes M et N sont étroitement liés avec le

VIScpz alors que les groupe O et P sont plus proches du VISgor, lui même apparenté au VIScpz[167,170,171] (Figure **13**).

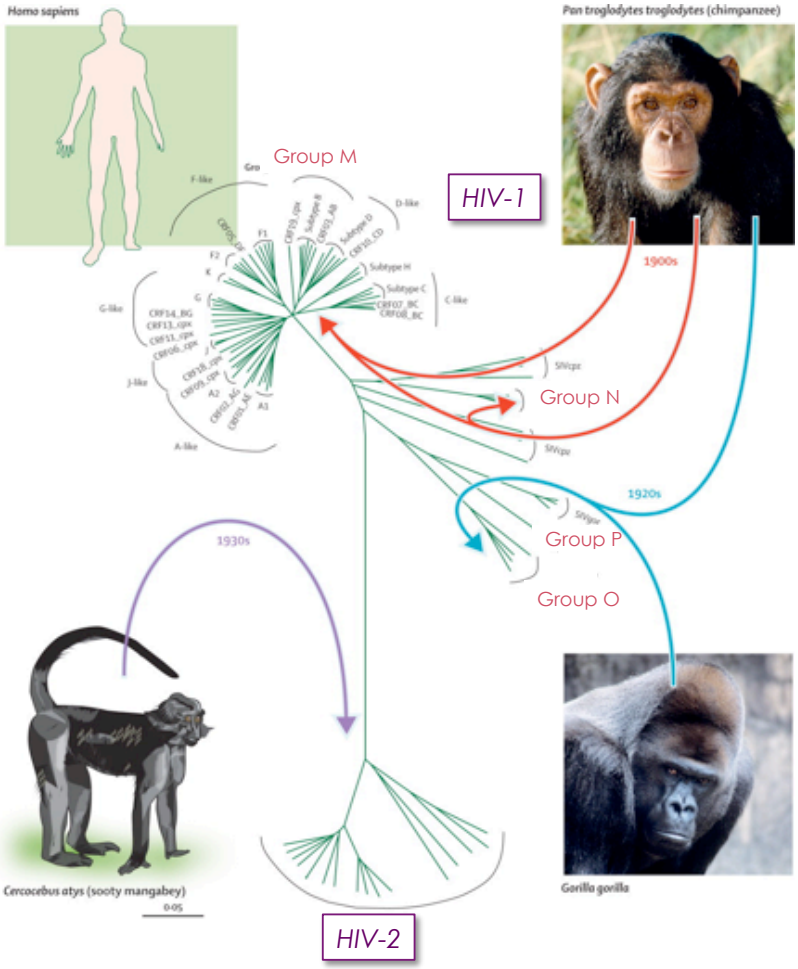
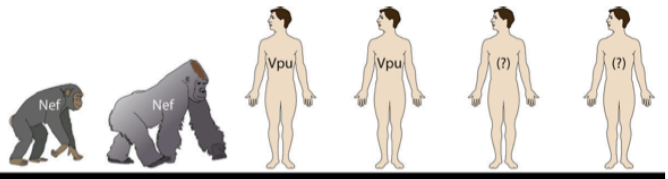


Figure 13 : Arbre phylogénétique du VIH. Deux types de VIH, le type 1, proche des virus simiens (VIS) des chimpanzés et le type 2 proche du VIS du mangabey enfumé. Les quatre groupes (M, N, O et P) du VIH-1 sont indiqués en rouge. Le VIS chimpanzé est à l'origine du VIS de gorille, lui même relié aux groupes P et O. Adapté de [172].

Le groupe M et le plus répandu et est responsable de la pandémie de SIDA (près de 37 millions de personnes infectées recensées en 2014). Il est lui-même subdivisé en neuf sous-types (A-D, F-H, J et K) et en formes recombinantes, les CRF (circulating recombinant forms) (voir en aval, ***variabilité apportée par la RT***). A l'inverse, les groupes N et P ont été caractérisés chez un nombre limité de patients (20 individus et 2 individus, respectivement). Cette restriction peut s'expliquer par une capacité infectieuse moindre (voir ci-dessous). Enfin, le groupe O est le plus courant, après le M, avec près de 100 000 individus infectés, mais ne s'est pas répandu géographiquement, la majorité des cas décrits étant localisés en Afrique[172,173] (Figure **13**).

La propagation mondiale du groupe M ainsi que sa prévalence sont probablement dues à l'adaptation optimale de ses gènes auxiliaires. En effet, comme décrit précédemment le VIH dispose de plusieurs protéines auxiliaires, qui n'ont pas toutes la même activité selon les groupes. Chez le VIH-1 groupe M, la protéine Vpu est responsable de la dégradation des CD4 dans la cellule et de l'inhibition de la restriction virale effectuée par Tetherine, alors que la protéine Nef est impliquée principalement dans la dégradation des CD4 à la surface de la cellule. De façon étonnante, chez les VIScpz et VISgor, l'antagoniste de Tetherine est la protéine Nef, alors que ce rôle est effectué par Vpu chez le VIH-1/ M, évolution qui a pu survenir lors du changement d'hôte ou lors de la spéciation. Le rôle de dégradation de CD4 de Vpu est conservé chez les VIS et le VIH-1 à l'exception du groupe N. Pour le VIH-1, seuls les groupes M et N utilisent Vpu pour contrer Tetherine. Au vu des données épidémiologiques, l'adaptation des protéines auxiliaires Vpu et Nef à l'hôte humain est corrélée avec le pouvoir infectieux des virus, puisque seul le groupe M possède des protéines totalement fonctionnelles. Les protéines Vpu du groupe N et Nef du groupe P sont inactives dans la dégradation de CD4 et l'antagonisme de Tetherine, respectivement, ce qui résulte en une faible capacité infectieuse et donc la propagation limitée des virus de ces groupes. Chez le groupe O, ni Vpu, ni Nef n'ont le rôle d'inhiber l'action antivirale de Tetherine, cependant, au vu de leur importance pandémique, il est probable que les virus du groupe O utilisent un autre facteur pour contrer Tetherine[174,175] (Figure **14).**



		SIVcpz	SIVgor	HIV-1 M	HIV-1 N	HIV-1 O	HIV-1 P
Nef	CD4	+	+	+	+	+	+
	Tetherin	+	+	–	–	–	–
Vpu	CD4	+	+	+	–	+	+
	Tetherin	–	–	+	+	–	–

Figure 14 : Evolution des rôles de Vpu et Nef. Ciblage de CD4 et Tetherine médiée par Vpu et/ou Nef selon la souche virale (VIS chimpanzé, VIS gorille, VIH-1/M, VIH-1/N, VIH-1/O et VIH-1/P). Adapté de [174].

b. Variabilité apportée par la RT

En plus de ses nombreuses origines, le VIH a une forte capacité de renouvellement génétique en partie due à la RT. En effet, la transcriptase inverse ne possède pas de correction d'épreuve et introduit en moyenne 1 à 3 mutations par génome et par cycle infectieux[176,177]. Ces erreurs de réplication associées au taux élevé de renouvellement des virions, qui est de plus de 10^{10} virions produits par jour chez un individu infecté, génèrent une variabilité accrue observée chez les patients infectés[178].

En plus de ces erreurs de réplications, la RT catalyse environ 3 à 4 évènements de recombinaison par génome et par cycle infectieux. La synthèse de l'ADN viral implique deux transferts de brins pour générer les LTRs, qui peuvent être intramoléculaires ou intermoléculaires[101]. De plus, la RT est capable de changer de matrice ARN durant la synthèse, menant à la formation d'un ADN recombinant si les deux ARN encapsidés sont génétiquement différents. En effet, lors de la coinfection par des virus différents, les néoparticules virales peuvent être hétérozygotes (encapsidation de deux ARN distincts). Lorsqu'une telle particule virale infecte une nouvelle cellule, la recombinaison lors de la transcription inverse permet le réassortiment des polymorphismes de chaque génome et l'émergence de nouveaux variants du virus[179,180] (Figure **15**).

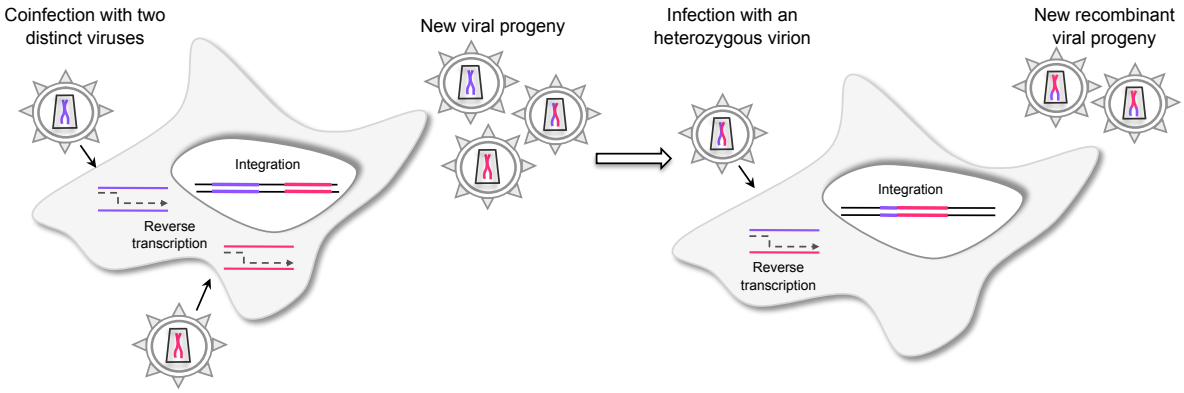


Figure 15 : Variabilité apportée par la recombinaison. La coinfection par deux virus différents (violet et rose) engendre la production d'une population virale hétérogène dont des virus dit hétérozygotes possèdent deux ARN encapsidés différents. Lors de l'infection d'une cellule par ce virus hétérozygote, le saut de brin lors de la transcription inverse génère un ADN recombinant, qui une fois intégré au génome cellulaire, amène à la production de virus recombinants. Adapté de [182].

Plusieurs modèles moléculaires ont été décrits pour expliquer le transfert de brin, qui se produirait préférentiellement lors de la synthèse du brin négatif[181]. Un premier modèle, appelé "forced copy choice", consiste au changement de brin par la RT dû à des lésions sur sa matrice ARN. En effet, la dégradation de l'ARN par le domaine RNaseH et la polymérisation de l'ADN par le domaine polymérase étant concomitante, il est possible que la RT change alternativement de matrices afin de synthétiser un ADN viral complet. Cette hypothèse est supportée par la découverte d'ARN génomiques discontinus qui mènent encore à la production de particules virales[182]. Un autre modèle, le "copy choice", suggère que la recombinaison est influencée par des facteurs viraux (NC, RNase H, structure de l'ARN)[183,184] mais nécessite, tout comme le "forced copy choice", une homologie de séquences, démontrée par la baisse du taux de recombinaison dans le cas de matrices non homologues[185].

c. *Impact de la recombinaison*

Le génome du VIH est composé de résidus conservés, important pour la fonctionnalité, et d'acides aminés non conservés originaire de sources de variabilité diverses (mutations apportée par la RT ou par A3G/F, origine simienne différentes, ...). Ces résidus variables prennent toute leur importance lors de la recombinaison, puisqu'ils sont redistribués lors de tels évènements, générant un ADN recombinant composé de résidus conservés et d'un nouveau sets d'acides aminés non conservées provenant des ARN parentaux.

La combinaison d'acides aminés variables peut permettre l'émergence de virus avec de nouvelles propriétés, lui apportant un avantage par rapport aux parentaux. En effet, elle peut permettre la combinaison de mutations avantageuses, comme par exemple l'assemblage de plusieurs mutations de résistances aux antirétroviraux, générant un nouveau type de virus plus résistant aux traitements[166]. Elle peut également favoriser l'élimination de mutations délétères en restaurant le génotype initial, ou encore la génération d'un ADN fonctionnel à partir d'ARNs parentaux endommagés (cassure de brin, bases modifiées, etc.) dans le cas du modèle "forced copy choice"[186].

En outre, la recombinaison fait partie intégrante du cycle de vie du virus et favorise une évolution rapide qui lui permet de s'adapter plus facilement à son hôte. Ces virus recombinants peuvent présenter des avantages par rapport aux virus parentaux et être sélectionnés. Lorsqu'un recombinant est identifié chez un patient on parle de forme recombinante unique (URF) et s'il s'établit dans la pandémie et infecte plusieurs hôtes, on parle alors de CRF. Les CRF sont nommés par un nombre (dans l'ordre croissant de leur découverte) suivi par les lettres des sous-types parentaux. Ces nouveaux virus représentent près de 20% des infections dans les régions où plusieurs sous-types différents co-circulent, montrant l'importance des évènements de recombinaison (plus de 30 CRF identifiées, Figure **13**) dans la propagation du groupe M à l'échelle mondiale[187]. De plus, malgré la grande distance génétique au niveau de la séquence nucléotidique observée entre les groupes M et O[188], plusieurs virus recombinants M/O ont émergé chez les patients[189-191]. La découverte de tels recombinants intergroupes provenant de virus phylogénétiquement éloignés démontre l'existence possible de bonds dans l'évolution qui peuvent avoir des conséquences encore mal évaluées sur les propriétés virales.

Cette diversification constante des séquences est cependant limitée par des contraintes coévolutives visant à préserver la fonctionnalité des protéines. En effet, ces évènements de saut de brin devraient se produire théoriquement tout le long du génome, avec des

préférences possibles pour certaines régions. Cependant, une précédente étude du laboratoire, basée sur l'analyse des points de cassures (site à partir duquel la séquence change d'origine) de plusieurs souches recombinantes, montre que la recombinaison se produit préférentiellement à des sites spécifiques, que l'ont appelle points chauds, alors que certaines zones du génome, les points froids, présentent très peu de points de cassures[192,193]. Ces observations suggèrent que la sélection naturelle influence le schéma de répartition des points de cassures. En effet, la présence de points froids au sein du génome signifie que les potentiels recombinants dans ces régions sont contre sélectionnés dans la pandémie, car probablement moins infectieux, démontrant l'importance des résidus non conservés (redistribués lors des évènements de recombinaison) dans le maintien de la fonctionnalité des protéines du virus[194,195].

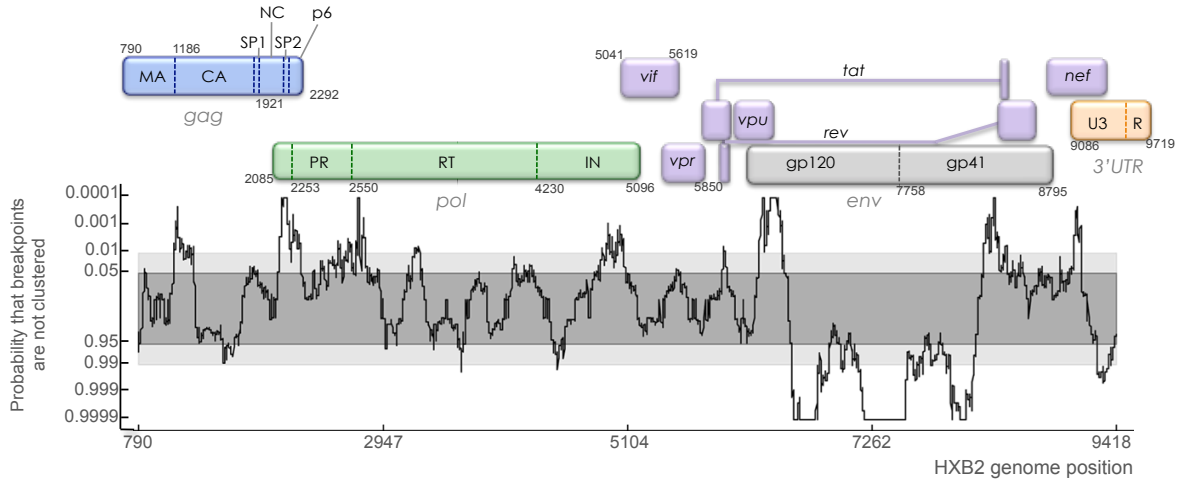


Figure 16 : Fréquence de distribution des points de cassures le long du génome du VIH-1/M. Panneau du dessus. Représentation graphique de la distribution des points de recombinaison le long du génome en fonction de la probabilité que celle ci n'est pas différente de celle attendue par hasard. La carte du génome de la souche HXB2 (HIV-1/M) est donnée comme référence. Adapté de[200].

En effet, la recombinaison peut rompre les liens génétiques entre différentes parties du génome qui sont connectées par coévolution. Comme le VIH présente une diversité génétique importante, on peut supposer que la fonctionnalité de ses protéines repose non seulement sur les résidus conservés, mais aussi sur les acides aminés non conservés, qui font partie de réseaux de coévolution permettant de contrebalancer l'effet potentiellement délétère d'une mutation par la sélection d'une ou plusieurs mutations compensatoires dans une autre position de la protéine. Ces mutations forment un réseau dynamique d'interactions nécessaire pour le maintien de l'activité de la protéine que l'on appelle réseaux de coévolution. Lors d'une recombinaison entre deux souches phylogénétiquement distantes, le réassortiment des polymorphismes peut perturber ces réseaux car les résidus non conservés n'auront pas la même histoire évolutive et peuvent donc être incompatibles, interférant ainsi avec la fonctionnalité[194,195].

d. *Etudes de la coévolution*

Les événements de coévolution sont probablement indispensables dans le maintien de la fonction et de la structure des protéines du VIH, afin d'éviter que la forte diversification génétique entraîne une inactivation du virus. Ces contraintes coévolutives définissent un carrefour entre l'hypervariabilité du virus et la fonctionnalité de ses protéines et comprendre leur impact phénotypique est important pour déterminer les limites de variabilité du virus et concevoir une thérapie plus efficace.

De nombreuses approches pour l'étude des réseaux de coévolution ont vu le jour, dont l'approche *in silico*. Plusieurs techniques consistant en l'analyse de données de séquençages par des modèles statistiques[196], ou en l'inférence directe de la fréquence d'apparition des mutations[197], ont permis l'identification de nombreux variants. Cependant, ces approches informatiques, bien que rapides et peu coûteuses, présentent des désavantages, tels que la probabilité de confondre une mutation avec une erreur de séquençage, le fort taux de faux positifs, la nécessité d'avoir une couverture des séquences maximales et une population très importantes pour limiter les erreurs et obtenir des données significatives[198].

D'autres approches, *in vivo*, utilisent la variabilité naturelle du VIH (recombinants, groupes phylogénétiques) pour l'étude des protéines dont les acides aminés covarient[199-201], avec une fiabilité supérieure aux techniques de bioinformatique bien que plus coûteuse en temps. Ces approches ont également permis d'identifier des réseaux de coévolution et de directement évaluer leur impact sur l'infectivité du virus. D'ailleurs, le laboratoire a mis en évidence, dans l'étude citée précédemment (Figure **16**), la faible répartition de points de recombinaison au sein du gène *env*, soulevant l'hypothèse que des réseaux de coévolution pourraient être impliqués dans le maintien de la fonctionnalité de la gp120 et de la gp41 (protéines codées par *env*). Une approche par construction et tests de protéines d'enveloppe chimères a permis de mettre en évidence un résidu non conservé de la région C2 de la gp120, qui, lorsqu'il est substitué par l'acide aminé naturellement présent dans une autre souche phylogénétique du VIH, abolit totalement l'entrée de la particule virale dans la cellule probablement en interférant avec le changement de conformation qui succède à la liaison au corécepteur (voir **I.3.a**). L'identification de mutations compensatoires au sein de la boucle V3 qui permettent de restaurer entièrement la perte de la fonctionnalité du mutant ponctuel de la région C2 a permis de mettre en évidence une relation coévolutive entre ces deux régions de la protéine, montrant l'importance des régions variables de l'enveloppe pour allier la diversification génétique de la protéine et la préservation de sa fonctionnalité.

II. L'intégrase

1. Structure de la protéine

Comme décrit précédemment, l'intégrase est composée de trois domaines structurés, le NTD, le CCD et le CTD reliés par des régions flexibles (Figure **17**). Chaque domaine est essentiel pour l'intégration de l'ADN viral dans le génome de la cellule hôte. La possibilité de restaurer la fonctionnalité en combinant deux intégrases, préalablement inactivées par mutations à des positions différentes dans les domaines l'IN a montré qu'elle agit sous forme de multimères[202], un dimère d'IN étant suffisant pour assurer le clivage de l'ADN viral[203]. La structure en résonnance magnétique nucléaire du NTD[204] et du CTD[205] et la structure cristalline du CCD[206] révèlent une organisation dimérique pour chacun des domaines.

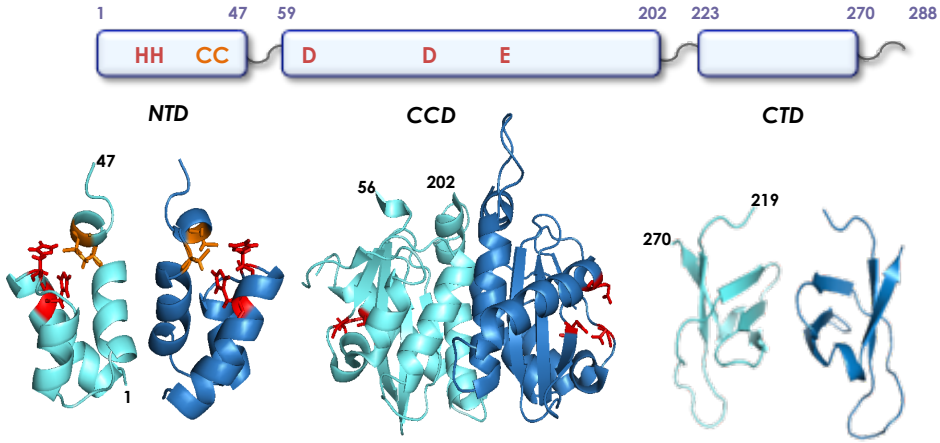


Figure 17 : Structure des domaines de l'intégrase du VIH-1/M. La partie supérieure est une représentation schématique des domaines et la partie inférieure contient les structures cristallographiques des domaines sous forme de dimères. Le NTD comporte un motif en doigt de zinc HHCC indiqué sur le schéma et reporté avec les mêmes couleurs sur la structure. Le CCD contient la triade catalytique DDE, indiquée en rouge sur le schéma et reportée sur la structure avec les mêmes couleurs. Adapté de [230].

a. Domaine N-terminal

Le domaine amino-terminal allant du résidu 1 à 46, composé d'un faisceau de trois hélices α , est structuré sous la forme d'un domaine "hélice-tour-hélice" (HTH) et comprend un motif de liaison au zinc conservé chez tous les rétrovirus[204] (Figure **17**). Ce motif en doigt de zinc est constitué de deux histidines (H_{12} et H_{16}) et de deux cystéines (C_{40} et C_{43})[207]. La liaison de l'ion métallique contribue à la multimérisation fonctionnelle de la protéine et à la stabilisation de sa structure. La mutation de ces résidus engendre une intégrase monomérique, incapable de catalyser l'intégration[208].

D'autres études ont montré que la substitution de la lysine en position 14 par une alanine (mutation $K_{14}A$) déstabilisait le tétramère d'intégrase et affectait également l'interaction de l'IN avec LEDGF[209], inhibant ainsi l'intégration. D'ailleurs, ce résidu a montré être à l'interface NTD-CCD entre deux dimères par son interaction avec le motif de multimérisation ($K_{186}R_{187}K_{188}$) du CCD, contribuant ainsi à l'interface d'interaction du tétramère d'IN. La lysine 188 a montré être peu impliquée dans la formation du tétramère puisque lorsqu'elle est mutée en alanine, l'activité n'est que peu perturbée, alors que l'IN est totalement inactive lorsque ce sont K_{186} ou R_{187} qui sont mutés[210].

Plusieurs résidus du NTD ont également été montrés importants pour l'infectivité (par exemple V_{32} A_{33} K_{34})[211,212]. Le NTD est relié au CCD par une région flexible allant du résidu 47 à 55, et plusieurs résidus de cette boucle ont suggéré être impliqués dans la liaison à l'ADN viral dans le complexe de transfert de brin de l'IN (Figure **18.B**)[58].

b. Domaine catalytique central

Le domaine catalytique central, allant du résidu 56 à 186, est composé de 6 hélices α et 5 feuillets β et son repliement général est similaire à celui des protéines de la superfamille des polynucléotidyles transférases, dont la RNase H fait partie[206,213].
Ce domaine contient la triade catalytique, $D_{64}D_{116}E_{152}$, conservée chez tous les Rétrovirus et les transposases, qui lie des ions métalliques divalents (magnésium, Mg_2^+ ou manganèse, Mn_2^+) afin d'initier l'intégration (Figure **17**).

Le CCD arbore de nombreux autres résidus essentiels à l'intégration. Tout d'abord, le domaine CCD est suggéré être impliqué dans la liaison à l'ADN. Plusieurs régions (49-69, 137-152, 153-167)[214] ou résidus ($V_{72}Y_{143}S_{147}Q_{148}S_{153}K_{159}K_{160}I_{161}G_{163}V_{165}H_{171}L_{172}$)[215,216] ont été identifiés par photocrosslinking ou par modélisation[58] (Figure **18.B**) pour la liaison à l'ADN viral. De même que pour la liaison à l'ADN cellulaire[217] pour laquelle plusieurs résidus ont également été identifiés[58,218] (Figure **18.B**).
Ce domaine contient également la surface d'interaction avec le cofacteur LEDGF ($A_{128}A_{129}W_{131}W_{132}I_{161}V_{165}R_{166}E_{170}L_{172}K_{173}$)[219,220], ainsi que le motif de multimérisation ($K_{186}R_{187}K_{188}$)[209,201,221,222] , nécessaire pour la formation du tétramère. D'ailleurs ce domaine contient également la surface d'interactions impliquée dans la stabilisation du dimère ($W_{61}E_{85}E_{87}K_{103}R_{107}W_{108}$)[223]. La boucle allant du résidu 187 à 194, qui relie les hélices α5 et α6, contient le résidu F_{185}, dont la mutation spécifique en lysine ($F_{185}K$) augmente considérablement la solubilité de la protéine mutante[224] qui a été utilisée pour obtenir les structures cristallines du domaine seul ou avec le NTD ou le CTD[225].

c. Domaine C-terminal

Le domaine carboxy-terminal, allant du résidu 212 à 288 pour les VIH-1/M et 298 pour les VIH-1/O est composé de 5 feuillets β, dont le repliement est similaire à la forme d'un domaine SH3 (sarc homology domain)[205] (Figure **17**). Ce domaine est très riche en résidus basiques et possède deux régions conservées chez les Lentivirus essentielles pour la réplication du VIH-1 ($_{235}$WKGPAKLLWKGEGAVV$_{250}$ et $_{259}$VVPRRK$_{264}$)[226]. La région 220-270 est responsable de la liaison à l'ADN[227,228] de manière aspécifique. D'ailleurs, plusieurs résidus dans ce domaine sont suggérés être impliqués dans des interactions avec l'ADN. La majorité des contacts entre le CTD et l'ADN implique des lysines ou des arginines (R_{228},R_{231},E_{246},A_{248},R_{263},K_{266})[58,229], suggérant des interactions électrostatiques avec les phosphates du squelette de l'ADN[229]. De plus, deux mutations de ce domaine (L_{241}A, L_{242}A) ont été identifiées pour perturber la dimérisation de l'IN[230], et donc, compromettre l'activité de l'enzyme, montrant que ces deux résidus sont probablement impliqués dans des contacts inter domaines permettant de stabiliser le dimère d'IN. Ce domaine contient également un résidu dont la mutation spécifique (C_{280}S) augmente la solubilité de la protéine mutante[224]. Enfin, ce domaine contient le site d'interaction avec la transcriptase inverse (voir **II.3.a** rôle dans la transcription inverse).

d. Intasome

Les études structurales de l'intégrase multimérique en complexe avec l'ADN, appelé aussi intasome, sont particulièrement difficiles au vu de la tendance à agréger de l'intégrase du VIH-1. Ces propriétés biochimiques défavorables nécessitent l'utilisation de mutations solubilisantes qui changent les propriétés de la protéine. Un modèle de cryo-électromicroscopie (cryo-EM) de l'intégrase du VIH-1/M en complexe avec l'ADN viral et cellulaire et son cofacteur LEDGF/p75 suggère que sa forme active est un dimère de dimères asymétriques et que chaque dimère possède une activité catalytique[231] (seul un monomère est actif au sein du dimère), en concordance avec la présence majoritaire de tétramère d'IN dans des extraits nucléaires de cellules infectées[232] et l'action de l'IN sous forme de dimère pour cliver l'ADN viral[203].

Comme les intégrases rétrovirales partagent une structure conservée en trois domaines, des modèles structuraux d'IN d'autres virus, homologues au VIH, ont été décrits, dont notamment l'intasome du virus foamy prototypique de la famille des spumavirus (PFV), obtenu par cristallographie[233,234]. La structure de l'intégrase de PFV montre une architecture tétramérique (comme le modèle de celle de l'IN de VIH-1/M), dont une partie centrale avec une structure conservée chez plusieurs autres rétrovirus. En effet, une structure centrale similaire a été décrite pour d'autres intégrases (RSV, MMTV, MMV) qui présentent pourtant

une structure multimérique différente. Les intégrases du virus du sarcome de Rous[235] (RSV) et du virus de la tumeur mammaire de la souris[236] (MMTV) présentent un arrangement octamérique, alors que l'intégrase du virus visna-maëdi[237] (MVV) s'organise en tétramère de tétramères (hexadécamère). Pour ces différentes intégrases, la structure centrale est conservée et est encadrée de protomères additionnels (4 pour RSV et MMTV et 12 pour MVV).

Récemment, une étude a démontré que la fusion de la protéine de liaison à l'ADN Ssod7 en N-ter de l'intégrase du VIH-1, améliorait sa solubilité, tout en conservant son activité *in vivo*. Le modèle hautement résolutif de cryo-EM de l'intégrase fusionnée à Ssod7 en complexe avec l'ADN viral et cellulaire, mimant le complexe de transfert de brin (STC), montre une forme tétramérique, comme le suggère le modèle de l'intasome en liaison avec LEDGF/p75[231], dont les interactions entre les domaines CCD et CTD des dimères asymétriques stabilisent la structure (Figure **18.A**)[58]. A partir de la structure en cryo-EM du STC du VIH-1 des prédictions d'interactions avec l'ADN viral ou cellulaire (Figure **18.B**), ainsi que des interactions inter domaines (Figure **18.C**) ont été faites.

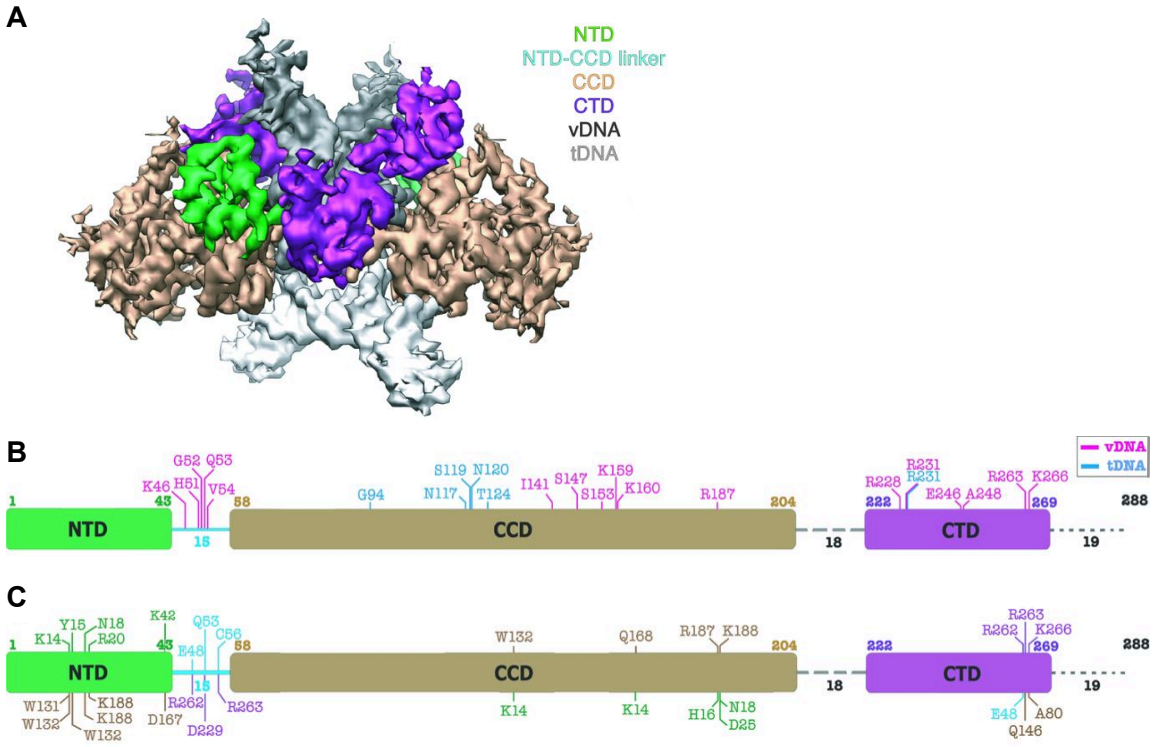


Figure 18 : Modèle du tétramère d'integrase en complexe avec l'ADN viral et cellulaire. A. Modèle obtenu par cryo-EM du dimère de dimères asymétriques de l'intégrase du VIH-1 en complexe avec l'ADN viral (gris foncé) et cellulaire (gris clair). Les protomères d'IN sont colorés en fonction des domaines : NTD (vert), boucle flexible reliant NTD au CCD (cyan), CCD (brun), CTD (violet). **B.** Carte de prédiction des résidus en contact avec l'ADN viral (rose) et cellulaire (bleu) au sein de l'intasome d'IN. Le code couleur des domaines est conservé. **C.** Carte de prédiction des interactions inter domaines au sein du complexe. Les résidus indiqués en dessous du schéma interagissent avec ceux indiqués, en face, au dessus du schéma. Le code couleur des domaines est conservé. Pour augmenter sa solubilité l'IN est fusionnée en N-terminal avec la protéine Sso7d. Adapté de [58].

Des données additionnelles ont été apportées par la reconstruction de la carte de densité électronique du STC en liaison avec le domaine de liaison à l'IN de LEDGF/p75, révélant une structure plus importante avec douze protomères d'IN additionnés au tétramère (formant ainsi un hexadécamère), très similaire à la structure de l'hexadécamère de MVV, appelée STC d'ordre supérieur. Dans ce modèle on retrouve la partie centrale, le STC tétramérique, conservée chez les intégrases d'autres Rétrovirus (PFV, RSV, MMTV, MVV), flanquées de protomères additionnels. Les domaines CTD sont réorganisés dans le STC d'ordre supérieur pour former une interface entre les protomères d'IN et plusieurs résidus sont présumés relevant pour des interactions dans le STC d'ordre supérieur et non dans le STC tétramérique (E_{35}, E_{212}, K_{219}, K_{240}, L_{242}, K_{244}, I_{257}, V_{259}, R_{269}).

2. Mécanisme d'action

L'intégration se déroule en deux réactions de trans-éstérification, médiées par l'intégrase et qui impliquent la cassure d'une liaison phosphodiester[238]. La première concerne l'ADN viral avec le clivage d'un dinucléotide en 3' des LTRs de l'ADN viral, la deuxième implique l'ADN cellulaire avec le transfert du brin d'ADN viral au génome de la celle hôte. Les nucléotides additionnels et les régions simples brins sont ensuite réparés par la machinerie cellulaire.

a. Clivage et transfert de brin

L'intégration débute par la formation de l'intasome, chaque dimère d'intégrases se lie à l'ADN viral au niveau de l'extrémité 3' des LTRs. Le dinucléotide qui succède au CA canonique, conservé chez tous les Rétrovirus, est retiré à chaque extrémité par les monomères d'integrase catalytiquement actifs au sein des dimères, générant deux extrémités 3'OH réactives, CA_{OH}[239,240] (Figure **19**). Le brin d'ADN viral clivé résultant est utilisé comme substrat pour le transfert au génome cellulaire, suite à l'import nucléaire du PIC. Les deux réactions sont spatialement indépendantes, le clivage en 3' se produit dans le cytoplasme alors que le transfert du brin d'ADN viral au génome hôte implique une localisation nucléaire.

Les extrémités réactives vont s'insérer covalemment dans le brin d'ADN au niveau du sillon majeur par l'attaque nucléophile d'une liaison phosphodiester effectuée par l'extrémité 3'OH réactive de l'ADN viral, générant un ADN recombinant contenant le provirus[241-243] (Figure **19**). Le transfert de brin se produit simultanément et au même site pour chacune des deux extrémités virales. Les sites de liaison sur les deux brins d'ADN cible sont séparés par cinq paires de bases (dans le cas du VIH), ce qui entraîne une duplication de cinq nucléotides flanquant le provirus[243].

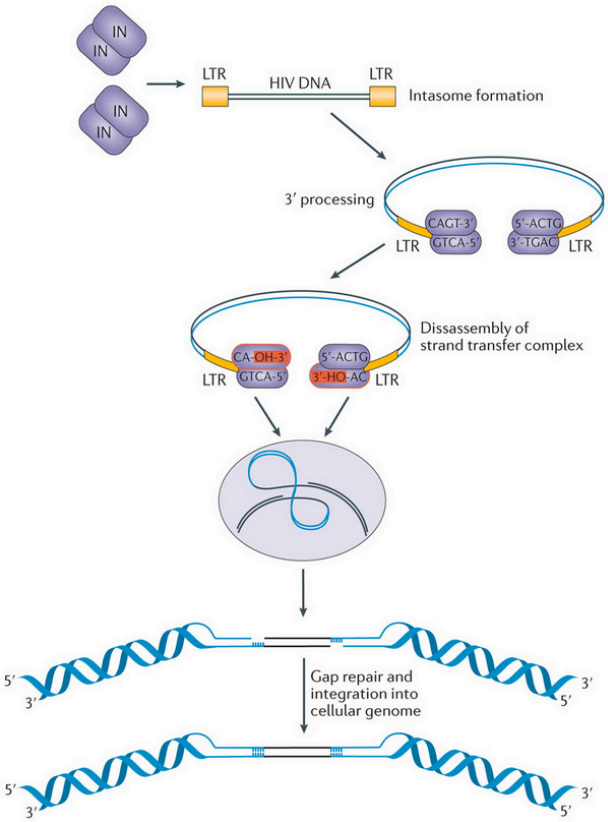


Figure 19 : Mécanisme d'intégration. Les dimères d'intégrases se lie au niveau des LTRs et forment l'intasome. Le dinucléotide GT en aval du CA à l'extrémité 3' des LTR est clivé, libérant une extrémité CA^{-OH} réactive, qui s'insère ensuite dans l'ADN cellulaire par transfert de brin. L'intégration est finalisée par la machinerie de réparation de l'ADN cellulaire. Adapté de [110].

L'intégration est complétée par la machinerie cellulaire de réparation de l'ADN qui va prendre en charge l'ADN ayant intégré le provirus. En effet, la coupure des dinucléotides d'origine virale aux extrémités 5' sortantes du provirus et la réparation des blancs dans la séquence d'ADN double brin cible sont nécessaires pour terminer la réaction globale d'intégration[244]. Les protéines Ku70 et Ku80 (sous-unité Ku), faisant partie du système de réparation de l'ADN par liaison des extrémités non homologues (NHEJ) ont montré être associées avec le PIC[245]. D'ailleurs, comme certaines formes non intégrées de l'ADN sont formées par le système NHEJ (cercles à 2LTRs, voir en aval, formes non intégrées de l'ADN viral), le rôle de la voie NHEJ dans la réparation de l'ADN après l'intégration a été suggéré[246], en accord avec des résultats précédents, montrant que l'intégration était réduite en l'absence de composés de cette voie de réparation (sous-unité Ku, kinase dépendante de l'ADN, X-ray repair cross-complementing protein 4)[247]. La réaction finale aboutit donc à l'insertion du provirus dans le génome hôte avec duplication d'une séquence chromosomique de cinq paires de bases de part et d'autre de l'ADN viral.

b. Choix du site d'intégration

En théorie, l'intégration peut se produire tout le long du génome, mais les analyses des sites d'intégration ont montré que ce procédé ne se fait pas au hasard et que l'intégrase cible préférentiellement les unités transcriptionnelles de l'euchromatine[248,249].

La chromatine, dont le premier niveau de compaction est la condensation de l'ADN en nucléosomes, existe sous deux formes dans le noyau, l'euchromatine et l'hétérochromatine. L'euchromatine, à l'inverse de l'hétérochromatine, est peu condensé et contient des gènes activement transcrits. Elle se caractérise également par des modifications post-traductionnelles (mono-, di- ou triméthylation, acétylation) des histones qui, complexés à l'ADN, forment les nucléosomes[250]. L'intégration a lieu préférentiellement dans l'ADN nucléosomal[251] due probablement à la préférence de l'intégrase pour l'ADN courbé, retrouvé à la surface des nucléosomes[252]. L'un des facteurs majeurs responsable du ciblage de l'intégration dans l'euchromatine est le cofacteur cellulaire LEDGF/p75. En effet, LEDGF/p75 agit comme un facteur de ciblage du site d'intégration via son interaction avec l'intégrase et son domaine d'interaction avec l'euchromatine[253,254]. En effet, LEDGF/p75 possède un domaine PWWP, qui lie spécifiquement les modifications d'histones associées à la transcription (comme par exemple la triméthylation de la lysine en position 36 de l'histone H3, $H3K_{36}me^3$) (voir en aval dans la partie **II.4.a** pour plus de détails).

D'autres facteurs cellulaires sont impliqués dans le ciblage de l'intégration. Des études de déplétion ont montré que CPSF6 favorise l'intégration dans les régions activement transcrites grâce à une localisation spécifique à la périphérie du noyau[255]. Cette localisation préférentielle s'explique par le rôle des protéines des pores nucléaire qui stabilisent LEDGF/p75 en périphérie nucléaire et régulent l'expansion de l'hétérochromatine à proximité du pore nucléaire[256,257].

Récemment, une étude a mis en évidence le rôle du complexe FACT (facilitating chromatine transcription) dans la stimulation de l'intégration du VIH-1[258] dans la chromatine. FACT est un complexe chaperon d'histones, hétérodimérique, composé de l'homologue humain du suppresseur de Ty16 (hSpt16) et de la protéine de reconnaissance de structures spécifiques (SSRP1). La sous-unité SSR1 de FACT possède un domaine HMG d'interaction spécifique avec le domaine PWWP de LEDGF/p75[259], suggérant que le complexe FACT pourrait faire partie du PIC grâce à son interaction avec LEDGF/p75, et moduler l'intégration. Comme FACT est associé avec la machinerie de transcription de l'ARN polymérase III, affectant ainsi la structure de la chromatine, il est supposé que ce complexe pourrait avoir un rôle dans le remodelage de la chromatine, assurant l'accès de l'intasome aux nucléosomes[258].

Ainsi, l'intégration de l'ADN viral a lieu préférentiellement dans les régions activement transcrites, à proximité des pores nucléaires, et est favorisée par plusieurs protéines cellulaires.

c. *Formes non intégrées de l'ADN viral*

Le mécanisme d'intégration a été longuement étudié et décrit comme n'étant pas totalement efficace : plusieurs produits secondaires de l'intégration peuvent être détectés durant l'infection[260]. En effet, l'ADN viral est retrouvé sous plusieurs formes non intégrées dans la cellule infectée, dont la forme majoritaire est l'ADN viral linéaire, substrat de l'intégration[261] (Figure **20**).

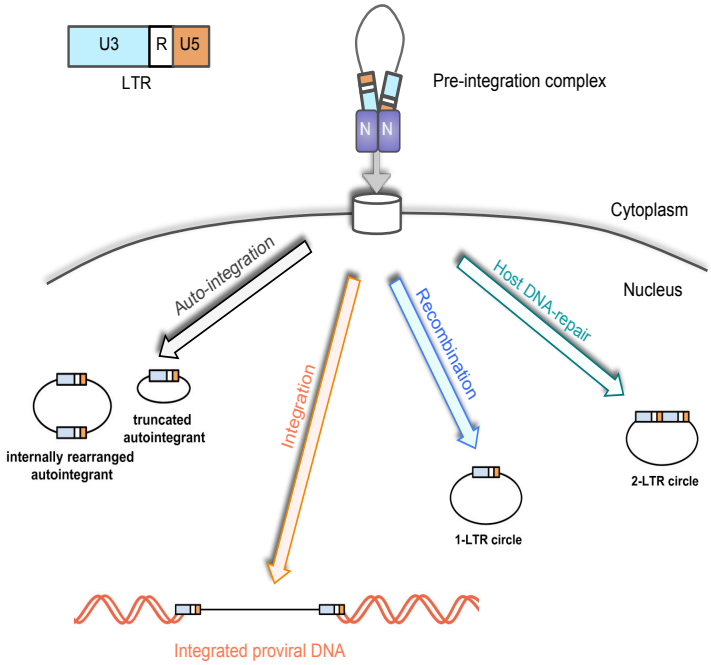


Figure 20 : Les différentes formes de l'ADN viral. L'autointégration peut conduire à la formation de formes circulaires tronquées ou réorganisées en interne. Des formes circulaires à 1LTR sont produites par recombinaison homologue. Les formes circulaires à 2LTRs sont formées par le système de liaison de jonctions non homologues (NHEJ).
Adapté de [260].

L'ADN viral existe également sous des formes circulaires qui peuvent prendre pour origine l'ADN linéaire complet ou l'ADN clivé aux extrémités 3' par l'IN. On trouve notamment des formes circulaires de tailles diverses, formées à partir de l'ADN viral clivé, et produites par l'autointégration, le brin d'ADN viral est transféré sur lui même par l'IN au lieu d'être inséré dans le génome cellulaire[262,263].

L'ADN viral est également retrouvé sous d'autres formes circulaires à 1LTR (1LTRc) ou 2LTR (2LTRc). Les cercles à 1LTR sont produits majoritairement dans le noyau par la recombinaison homologue de l'ADN linéaire sur les séquences répétées LTRs, alors que les cercles à 2LTRs sont produits exclusivement dans le noyau et sont classiquement utilisés dans la littérature comme marqueurs de l'import nucléaire.
Les 2LTRc peuvent être formés à partir de l'ADN viral, donnant lieu à des jonctions

palindromiques parfaites, résultant de la ligation des extrémités franches par le système cellulaire de jonction des extrémités non homologues (NHEJ), ou à des jonctions palindromiques imparfaites, plus généralement due à l'autointégration[264].
La quantité des cercles à 1LTR peut atteindre jusqu'à 30% de l'ADN viral total chez un virus infectieux alors que le taux de cercles à 2LTR est de l'ordre de 2 à 5% de la quantité totale d'ADN viral. Cependant, la relative abondance de ces différentes formes est dynamique et est dépendante des conditions d'infection virale. Par exemple, si l'IN est non fonctionnelle, la quantité de formes d'ADN viral non intégré va augmenter, de plus, si l'IN présente un défaut dans le clivage des LTRs, les 2LTRc formés à partir de l'ADN non clivé seront plus nombreux. D'un autre côté, dans le cas d'un défaut dans l'import nucléaire, il y aura une plus grande proportion de 1LTRc et, à l'inverse, une diminution des 2LTRc. Ainsi, la quantification *in vivo* des proportions relatives de ces différentes formes peut être informative de l'efficacité de ces étapes (import nucléaire, intégration)[265,266].

3. Activités non catalytiques de l'intégrase

Les mutations de l'intégrase peuvent avoir des effets multiples sur le cycle infectieux. Selon leurs effets, on distingue deux classes de mutants, la classe I, pour les mutations qui affectent l'intégration proprement dite et la classe II, pour celles qui n'affectent pas la catalyse directement mais d'autres activités de l'IN importante pour le cycle infectieux[267,268].
Plusieurs études de mutagénèse ont mis en évidence des mutations de l'intégrase ayant des effets délétères sur les étapes de la phase précoce (décapsidation, transcription inverse et import nucléaire) et de la phase tardive du cycle de réplication (encapsidation, maturation, morphogénèse)[269].

a. Rôle dans les étapes précoces du cycle de réplication

Les variations de la stabilité du cône de capside ont une influence sur la réplication, c'est pourquoi la **décapsidation** est une étape importante du cycle. Une étude a montrée que l'intégrase était requise pour le bon déroulement de cette étape. En effet, la délétion de l'IN affecte l'incorporation de la cyclophiline A (CypA) dans la capside qui, comme décrit dans la partie **I.3.a**, est importante pour l'infectivité[270]. L'IN est requise pour l'interaction entre la capside et la CypA, qui peut empêcher la décapsidation précoce menant à une perte d'infectivité.

La **transcription inverse** consiste en la conversion de l'ARN génomique viral en ADN, elle se produit au sein du complexe nucléoprotéique de transcription inverse, dérivé du cône de capside et qui contient plusieurs protéines virales (Vpr, IN, RT, CA, NC). Une étude a

démontré que certaines mutations de l'intégrase du VIH-1 induisent une diminution de la synthèse d'ADN plutôt qu'un défaut d'intégration. De plus, l'apport en *trans* d'une protéine de fusion IN-Vpr dans les virus mutants permettait la restauration de la transcription inverse par complémentation. Cependant, seul l'IN d'origine VIH-1 est capable de compenser les mutations et rétablir la synthèse d'ADN, la complémentation avec une IN d'origine VIH-2 n'a permis aucune restauration[271], démontrant une incompatibilité entre la RT VIH-1 et l'IN VIH-2. L'interaction physique et fonctionnelle entre la RT et l'IN a été mise en évidence *in vitro* par des expériences de GST-pulldown et de résonnance plasmonique de surface et *in vivo* par mutagénèse dirigée[271-273]. Cette surface d'interaction avec la RT a été identifiée dans le domaine C-terminal de l'IN (résidus $R_{231}W_{243}G_{247}A_{248}V_{250}I_{251}K_{258}$)[273,274] et concerne deux régions de la RT (les sous-domaine doigts et paume, résidu 1-242, et la partie C-ter du sous-domaine connexion, résidu 387-422)[272]. Ces régions de la RT faisant partie de la p66 et de la p51, l'IN pourrait se lier aux deux sous-unités, mais la proximité spatiale de ces deux régions de la p51 dans la structure de l'hétérodimère suggère que l'IN lierait la RT au niveau de la p51[272]. Cette interaction permettrait d'augmenter l'efficacité d'initiation de la transcription inverse et la procéssivité de la RT par sa stabilisation sur le complexe matrice-amorce[275].

Suite à la conversion de l'ARN en ADN, le complexe nucléoprotéique est appelé complexe de préintégration (PIC). Pour permettre la réplication du virus l'ADN viral doit être intégré au génome de la cellule, d'ou l'étape d**'import nucléaire** du PIC. Plusieurs études ont démontré l'importance de l'IN dans cette étape, par l'identification de possibles signaux de localisation nucléaire[276-278] (NLS) ou la mise en évidence de l'interaction fonctionnelle avec des facteurs cellulaires qui favorisent le passage actif du PIC par les pores nucléaires, tels que l'Importine 7[279], l'Importine 3[280] ou la Transportine 2[281], ainsi que la présence d'un NLS sur l'un de ses cofacteurs, le LEDGF/p75 (voir ci-dessous, **II.4.a**).

b. Rôle dans les étapes tardives du cycle de réplication

Après largage des particules virales, celles-ci nécessitent une étape de **maturation** pour devenir infectieuses. Les précurseurs Pr55Gag et Pr160Gag-Pol sont clivés par la protéase virale, faisant partie de Pol. La protéase s'active grâce à sa dimérisation au sein du précurseur dont l'IN fait partie. Le mécanisme par lequel l'IN peut influencer cette étape est encore méconnu, mais des études ont montré que la délétion de la séquence de l'intégrase provoquait un défaut de clivage des précurseurs[268,282,283], probablement à cause d'une mauvaise dimérisation des Pr160Gag-Pol, qui inhibe l'activation de la protéase. D'autre part certaines mutations ou délétions dans le C-terminal du domaine IN du précurseur Gag-Pol ont montré être responsables du défaut d'incorporation et de maturation du précurseur dans le virus bourgeonnant[268,282,283].

Lors de la maturation, le clivage des précurseurs en protéines individuelles permet la formation du complexe ribonucléoprotéique (RNP), entouré par le cône de capside. La bonne **morphogénèse** de la particule est tout aussi nécessaire à l'infectivité que la maturation. La RNP est composée des deux copies d'ARN génomique, d'ARN et protéines cellulaires et de protéines virales telles que l'IN, la RT et la NC. Une étude récente a montré que l'intégrase interagit avec l'ARN génomique virale et que cette interaction est impliquée dans l'assemblage correct du noyau viral contenant le complexe RNP pendant les derniers stades de réplication[285]. Une analyse par microscopie électronique de virus mutants, défaillants dans cette interaction, a montré qu'elle est essentielle à l'incorporation de la RNP dans le cône de capside, sa disruption par mutagénèse dirigée engendre une morphogénèse incorrecte des particules virales (ARN à l'extérieur du cône de capside)[284,285]. Ainsi, l'IN contribuerait à l'architecture et à la bonne localisation de la RNP dans la capside au sein de la particule virale mature.

4. Cofacteurs cellulaires

Comme d'autres protéines virales, l'intégrase s'associe à des facteurs cellulaires pour accomplir ses tâches au sein du complexe de pré-intégration. Les cofacteurs cellulaires de l'intégration ont été identifiés par différentes techniques (co-immunoprécipitation, test en double hybride sur Levures, test de restauration enzymatique de PIC inactivés par des sels avec des extraits cellulaires)[286]. Ces cofacteurs sont différenciés en deux classes, ceux qui s'associent au PIC de part leur affinité pour l'ADN, comme BAF et HMGA1, et ceux qui interagissent directement avec l'intégrase, comme INI1 et LEDGF/p75.

a. LEDGF/p75

Le cofacteur **LEDGF/p75** (Lens Epithelium Derived Growth Factor) fait partie de la famille des facteurs HDGF (Hepatoma Derived Growth Factor) et est retrouvé sous deux formes. L'une de 75kDa (p75) et l'autre, plus petite de 52kDa (p52) qui possède la même activité de régulation de la transcription mais ne lie pas l'IN du VIH, contrairement à l'isoforme p75.

En effet, LEDGF est composé de plusieurs domaines fonctionnels dont un domaine d'interaction spécifique avec les intégrases lentivirales[287] situé en C-terminal, absent chez l'isoforme p52. La structure de LEDGF/p75 en complexe avec l'IN a montré que son site d'interaction se situe dans une poche formée par la surface d'interaction entre les domaines catalytiques d'un dimère d'intégrase[219], située entre les positions 102-178. Plusieurs résidus ont été testés par mutagénèse ($W_{131}W_{132}I_{161}V_{165}R_{166}E_{170}L_{172}K_{173}$), leur mutation ne perturbent

pas l'activité in vitro de l'intégrase mais engendre des virus non infectieux, démontrant l'importance de LEDGF pour le cycle infectieux[219,220,288].

D'autres motifs fonctionnels ont été identifiés en N-terminal, des motifs de liaison à l'ADN, le motif PWWP et un signal de localisation nucléaire (NLS). Le NLS de LEDGF est composé de résidus basiques et présente des similarités avec le NLS du virus SV40. Une seule mutation dans le NLS abolit la localisation nucléaire de l'IN démontrant que LEDGF joue un rôle dans l'import nucléaire du PIC[289].

De plus, le facteur LEDGF, de part ses domaines de liaisons à l'ADN, stimule l'interaction de l'IN avec l'ADN et favorise le ciblage du site d'intégration dans les régions activement transcrites grâce au motif PWWP. Ce motif permet l'interaction spécifique avec les modifications d'histones, caractéristique de l'euchromatine[290,291].

b. Autres cofacteurs (BAF, HMGA1 et INI1)

La protéine cellulaire **BAF** (Barrier to Autointegration Factor) est essentielle et conservée chez les Métazoaires. Ce facteur de 89 acides aminés est sous forme de dimère en solution et stimule l'intégration intermoléculaire au sein du PIC *in vitro*[292,293]. Cette protéine ne stimule pas directement l'intégration, mais de part sa forte affinité pour l'ADN double brin et les protéines nucléaires, il est supposé qu'elle empêche l'autointégration grâce à la structuration spécifique de l'ADN viral[294].

Le cofacteur **HMGA1** (High Mobility Group chromosomal protein A1), anciennement nommé HMGI(Y), est une protéine de liaison à l'ADN qui peut moduler la régulation de la transcription et la structure de la chromatine. Cette protéine est organisée en trois domaines de liaison à l'ADN, séparés par des régions flexibles. Elle a été démontrée importante pour le fonctionnement du PIC *in vitro* et l'on suppose qu'elle joue un rôle dans l'intégration en compactant l'ADN dans une structure favorable au transfert de brin[295,296]. HMGA1 possède plusieurs sites de haute affinité pour la liaison de l'ADN viral dont plusieurs situés dans la région 5'LTR. Parmi ces différents sites de liaison, certains chevauchent des sites de liaison aux facteurs de transcription, comme par exemple AP1, suggérant que HMGA1 joue également un rôle dans la régulation de la transcription[297].

Le cofacteur **INI1** (Integrase Interactor 1) est le premier cofacteur à avoir été découvert par un test en double hybride chez la Levure[298]. En effet, INI1 est homologue à un facteur de transcription chez la Levure, SNF5. Cette protéine interagit avec l'IN (d'ou son nom) et est un composant du complexe de remodelage de la chromatine SWI/SNF. Ce complexe est connu

pour activer la transcription par le remodelage de la chromatine, soulevant la possibilité que INI1 permet le ciblage du PIC sur les régions ouvertes de la chromatine.

L'étude de la protéine a permis d'identifier deux régions très conservées, qui sont des répétitions imparfaites l'une de l'autre. La région répétée 1 est suffisante pour l'interaction avec l'intégrase, suggérant que seul la répétition 1 est impliquée dans l'interaction interprotéique alors que la région 2 pourrait avoir divergé pour assurer d'autres fonctions[299].

La structure du tétramère d'IN en complexe avec l'ADN viral et cellulaire, LEDGF et INI1 montre que INI1 stabilise l'IN dans une conformation incompatible avec l'intégration. De plus, INI1 siège dans le site de liaison de l'ADN cible, cette compétition pour la liaison de l'ADN cellulaire permettrait d'inhiber l'intégration non spécifique[300] (Figure **21**).

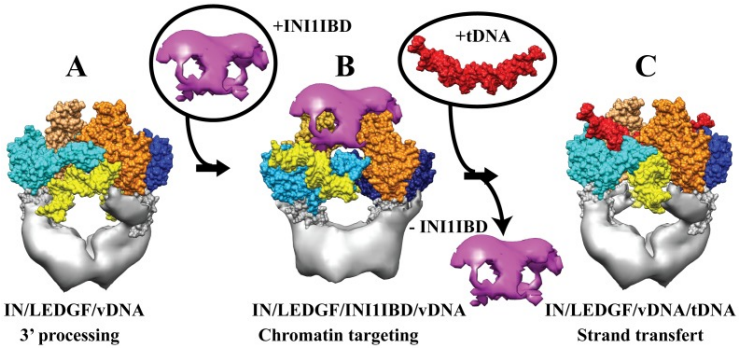


Figure 21 : Modèle du rôle de INI1. A. Structure du complexe de l'IN (tétramère schématisé en nuances de bleu de orange) avec LEDGF (représenté en gris) et l'ADN viral (en jaune). **B.** La liaison du domaine de liaison à l'IN de INI1 (INI1IBD, en violet) à la surface du complexe bloque l'IN dans une conformation intermédiaire. **C.** La libération de INI1IBD permet la transition dans la conformation pour le transfert de brin avec la liaison de l'ADN cible (en rouge). Adapté de [300].

5. Inhibiteurs de l'intégration

L'intégration est une étape clé du cycle infectieux et, à ce titre, une cible privilégiée pour le développement d'antirétroviraux. Plusieurs molécules sont en phase clinique ou déjà sur le marché pharmaceutique et l'on distingue deux classes principales, les inhibiteurs catalytiques, qui bloquent directement l'activité catalytique de l'intégrase, et les inhibiteurs non catalytiques ou allostériques.

a. Inhibiteurs catalytiques de l'intégrase

Deux types de molécules bloquant l'intégrase ont initialement été mis au point, les inhibiteurs de la liaison à l'ADN (INBI) et les inhibiteurs de transfert de brin (INSTI), qui ciblent l'intasome. Seuls les molécules de type INSTI ont été développé en phase clinique pour le traitement de patients. A ce jour, trois INSTI sont approuvés par la FDA, Raltegravir (RAL), mis sur le marché en 2007, suivi de Elvitegravir (EVG) en 2012, puis de Dolutegravir (DTG) en 2013[301,302].

Ces molécules partagent une structure commune constituée de deux ligands reliés par un segment, dont chacun possède une activité propre pour inhiber l'intégration. En effet, le premier ligand contient des atomes d'oxygène qui privent l'intégrase des cations métalliques, essentiels à son activité, en les séquestrant dans des complexes de chélation, qui sont par la suite éliminés par l'organisme. Le deuxième ligand contient un groupe aromatique benzyle halogéné qui interagit avec l'extrémité clivée de l'ADN viral, induisant ainsi son déplacement du site actif, ce qui empêche l'attaque nucléophile sur l'ADN cible et donc le transfert de brin[303].

RAL et EVG sont des inhibiteurs de transfert de brin dit de première génération et l'émergence de nombreuses mutations de résistances à ces inhibiteurs a nécessité le développement d'INSTI de seconde génération, dont DTG fait partie. DTG possède la même base commune constituée de deux ligands, mais est caractérisé par une structure plus rectiligne, comparé à RAL et EVG, qui lui confère une interaction plus stable avec l'intasome. En effet, le demi temps de dissociation de DTG de l'intasome est de 71h, comparé à 8,8h et 2,7h pour RAL et EVG, respectivement[301,304]. Cette meilleure association à l'intasome permet à DTG d'être efficace face aux souches de VIH résistantes aux INSTI de première génération (RAL et EVG).
En effet, suite à leur utilisation chez les patients, plusieurs voies de résistances (mutations qui, sélectionnées ensemble, rendent l'IN résistante aux inhibiteurs) à RAL et EVG ont été identifiées dont les principales mutations sont Y143C, Q148H et N155H de l'intégrase[304,305].

b. Inhibiteurs non catalytiques de l'intégrase

Les inhibiteurs catalytiques de l'intégrase ayant l'inconvénient de favoriser l'émergence de mutations de résistances, d'autres cibles ont été explorées afin d'inhiber l'activité de l'intégrase. Ainsi, plusieurs autres molécules ont été testées, dont les inhibiteurs non catalytiques de l'intégrase, désignés sous le nom d'inhibiteurs allostériques de l'IN (ALLINI).
La liaison de l'intégrase avec le cofacteur LEDGF/p75 ayant été décrite comme essentielle au bon déroulement du cycle infectieux[306], des ALLINI ciblant l'interaction IN-LEDGF/p75 (appelé aussi LEDGIN) ont été développés. Une autre étude a montré que l'inhibition de l'action de ciblage de la chromatine de LEDGF/p75, par surexpression d'une protéine LEDGF tronquée au niveau du motif PWWP dans des cellules infectées, est suffisante pour inhiber la réplication virale à des taux indétectables[307]. Par cette preuve de concept, plusieurs molécules qui ciblent spécifiquement l'interaction entre l'intégrase et LEDGF/p75, les LEDGIN, sont à l'étude. Ces différentes molécules ont un trait commun, un acide acétique qui mime le résidu D_{336} de LEDGF/p75, impliqué dans l'interaction avec l'intégrase,

concurrençant ainsi le cofacteur LEDGF/p75 pour la liaison à l'IN[308,309].

D'autres types d'ALLINI, qui ciblent d'autres interactions, sont également à l'étude. Par exemple, comme la dimérisation de l'intégrase est essentielle pour la réplication et concerne des résidus conservés chez les Lentivirus ($W_{61}E_{85}E_{87}K_{103}R_{107}W_{108}$)[223], des molécules ciblant l'interface de dimérisation formée par les CCD ont été développées[310]. Parmi celles ci, les peptides INH1 et INH5 ciblent une région du CCD de l'intégrase située entre les hélices α1 et α5. L'efficacité d'inhibition de ces peptides a été testée *in vitro*, et montre qu'ils inhibent de façon allostérique la formation du dimère d'intégrase et la catalyse[311].
Un mécanisme alternatif à l'inhibition de la formation du dimère a également été décrit, avec l'étude de molécules se liant à l'interface du dimère de CCD de l'IN du VIH-1, dans le but de stabiliser les interactions inter domaines dans une forme multimérique inactive. Un exemple de molécule, appelé compound 1, favorise la formation d'un multimère d'IN inactif, en acétylant le résidu K_{173}, et en se liant aux chaines latérales des acides aminés présents à l'interface des dimères de CCD ($E_{87}E_{96}T_{99}K_{103}$)[312].

III. Objectifs de l'étude

Le VIH est caractérisé par une diversité génétique importante, comme évoqué précédemment, et cette diversification de séquence peut interférer avec la fonctionnalité de ses protéines. En effet, les interactions inter- et intraprotéiques, nécessaires au maintien de la fonctionnalité des protéines, peuvent être brisées par l'apparition de mutations. Les évènements de coévolution permettent de compenser de telles mutations délétères. En effet, la coévolution des résidus d'une protéine signifie que ses acides aminés évoluent en parallèle, car chacun exerce une pression sélective sur l'autre, affectant ainsi son évolution, dans le but de conserver leur interaction. Ces résidus forment un réseau dynamique d'interactions nécessaire pour le maintien de l'activité de la protéine, et ne sont pas conservés au sein de souches phylogénétiquement distantes qui ont évolué indépendamment, car les réseaux n'ont pas forcément suivi les mêmes évènements évolutifs, ni subit les mêmes pressions de sélection.

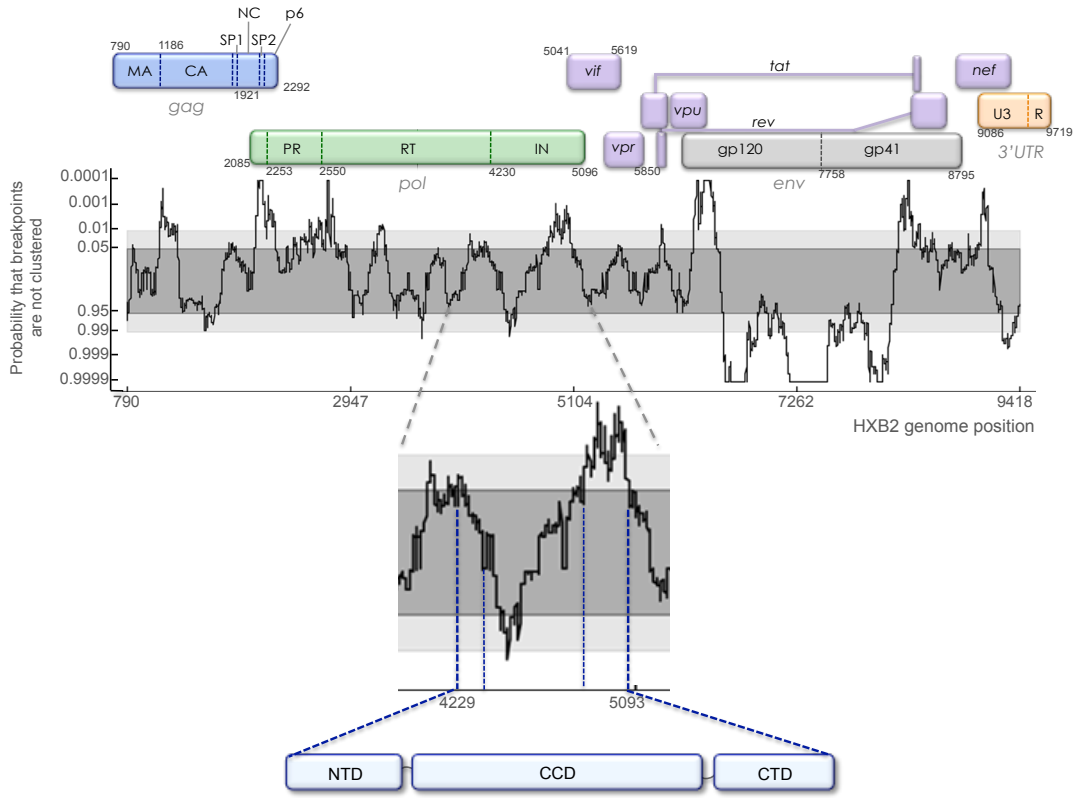


Figure 22 : Fréquence de distribution des points de cassures le long du génome du VIH-1/M. Panneau du dessus. Représentation graphique de la distribution des points de recombinaison le long du génome en fonction de la probabilité que celle ci n'est pas différente de celle attendue par hasard. La carte du génome de la souche HXB2 (HIV-1/M) est donnée comme référence. **Panneau du dessous.** Agrandissement de la région du génome codant l'intégrase. Les bordures des domaines de la protéine sont représentées par des traits bleus. Adapté de[200].

Un phénomène qui perturbe de façon importante les réseaux de coévolution est la recombinaison génétique. Au cours d'un cycle réplicatif, un évènement de recombinaison entre deux ARN différents introduit, contrairement à une mutation ponctuelle, plusieurs polymorphismes génétiques à la fois, rendant l'introduction de mutations compensatoires moins probable, ce qui peut perturber les réseaux de coévolution. En ce sens, une précédente étude du laboratoire, basée sur l'analyse de la répartition des points de recombinaison le long du génome du VIH-1, a montré que peu de points de recombinaisons étaient retrouvés dans le gène *pol*[200] (Figure **22**). Concernant l'IN, très peu d'évènements de recombinaisons sont retrouvés au centre de la portion de gène codant la protéine, suggérant que la plupart des recombinaisons qui ont lieu au sein de l'IN ne sont pas sélectionnées dans la pandémie. D'ailleurs, plus le point de recombinaison dans un gène est central, plus la probabilité de perturber les réseaux de coévolution par un événement de recombinaison est élevée et la plupart des variants dans l'IN sont probablement moins infectieux dû à la perturbation des réseaux de coévolution, phénomène attendu pour une protéine comme l'IN, qui interagit avec autant de partenaires et prend part à différentes étapes du cycle infectieux.

Comprendre les relations coévolutives entre différentes parties de la protéine est important pour définir les relations structure-fonction entre domaines dans les différentes phases du cycle réplicatif. Les études structurelles et fonctionnelles sur l'intégrase du VIH-1 ont été principalement réalisées à l'aide de mutations de résidus conservés. Mon objectif de thèse, en revanche, a consisté en la caractérisation des régions de l'intégrase impliquées dans la coévolution et d'identifier, pour chaque région, les acides aminés concernés ainsi que la nature du défaut fonctionnel. A cette fin, j'ai exploité la diversité génétique existante entre les groupes M et O du VIH-1. Ces groupes prennent pour origine deux virus de singe différents et présentent le degré de variabilité génétique le plus élevé parmi les groupes du VIH-1 (18% de variabilité entre M et O)[313], pourtant ce sont les seuls groupes, à ce jour, qui ont formé des recombinants qui ont émergé dans la pandémie.

Notre approche consiste en la génération d'intégrases chimériques entre des isolats primaires, afin d'analyser leur fonctionnalité. La construction de chimères permet le remplacement des acides aminés d'une intégrase (VIH-1/M) par ceux présents à la même position dans l'autre intégrase choisie (VIH-1/O). Ainsi, si les réseaux de coévolution spécifiques à chaque groupe sont perturbés dans les chimères, l'infectivité des virus portant ces protéines pourrait être altérée, reproduisant de façon similaire ce qui se produit naturellement avec la recombinaison génétique. La caractérisation du défaut fonctionnel et la cartographie des régions participant au réseau de coévolution seront informatives de l'implication de ces régions dans le cycle infectieux.

Matériels et Méthodes

1. Plasmides et construction des souches parentales

Pour étudier la fonctionnalité des protéines Pol du VIH-1, nous avons utilisé une variante modifiée du plasmide de transcomplémentation **p8.91CMVΔR,** décrit dans la littérature[314] et désigné ici par **p8.91** pour plus de facilité. Ce plasmide contient les gènes *gag* et *pol* de la souche de référence de laboratoire HXB2 (VIH-1/M, sous-type B). Les sites de restriction NotI et MluI ont été insérés en amont du gène *gag* et 18 nt en aval du début de la séquence codant la RT, respectivement. Cette insertion conduit à des modifications des acides aminés de la RT qui n'affectent pas sa fonctionnalité (E_6T, T_7R et $A_{554}S$). Avec le site de restriction SalI, déjà présent à la fin du gène *pol*, ces différents sites de restriction définissent deux cassettes de clonage : l'une englobant la séquence codante de la RT (MluI-RT-BspEI, 1680 pb), appelée par la suite RT, et l'autre englobant la séquence codant pour l'IN (BspEI-IN-SalI, 1561 pb), appelée par la suite IN. Les séquences RT et IN des deux isolats primaires utilisés pour cette étude ont été amplifiées et clonées dans p8.91 entre les sites MluI et SalI pour générer les plasmides parentaux de transcomplémentation correspondants. Pour faciliter le clonage des séquences IN (chimériques ou mutantes), nous avons utilisé les sites de restriction BspEI et SalI (Figure **23**).

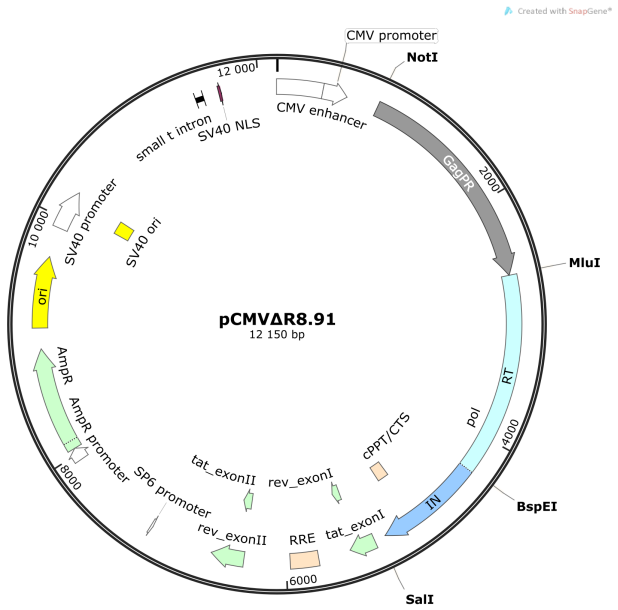


Figure 23 : Carte du plasmide de transcomplémentation p8.91. Les séquences d'origine HXB2 sont représentées en gris. Les cassettes RT, représentée en cyan et IN, schématisée en bleu, sont encadrées par les sites de restriction MluI, BspEI et SalI. La carte a été effectuée grâce au logiciel SnapGene.

Deux plasmides ont été construits pour l'amplification des courbes standard dans les différents essais de qPCR. L'un, appelé **pJet-1LTR**, pour la détection des produits tardifs et précoces de la transcription inverse a été obtenu en insérant la séquence du LTR (U3-R-U5) et la région Ψ (Psi) du VIH-1 provenant du pSDY (vecteur lentiviral décrit précédemment pour l'évolution dirigée des gènes cellulaires par le VIH [315]) dans le plasmide pJet avec le kit pJetPCRcloning (Thermo scientifique) (Figure **24**). Le deuxième, **pGenuine2LTR**, possède la jonction parfaite (non clivée) U5-U3 et a été construit par Eurofins Genomics (Allemagne).

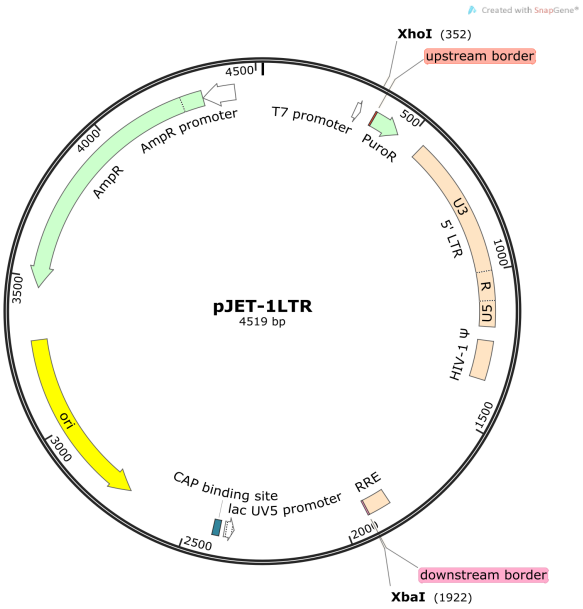


Figure 24 : Carte du plasmide pJet-1LTR. La séquence du 5'LTR (U3, R et U5) ainsi que la séquence ψ provenant du pSDY non modifié sont indiquées en orange, à l'intérieur de la cassette XhoI-XbaI. La carte a été effectuée grâce au logiciel SnapGene.

Pour produire l'ARN génomique des vecteurs lentiviraux, nous avons modifié le vecteur lentiviral pSDY (**pSDY**[315]), dans lequel le gène dCK a été remplacé par celui codant pour la RPF, afin de contrôler l'efficacité de la transfection par microscopie à fluorescence, et la séquence U3 du VIH-1 a été remplacée par celle de RSV dans le LTR en 5'(pSDY-RSV-RFP, désignée ici par pSDY pour plus de facilité).

2. Construction des intégrases chimères et mutantes

Les intégrases chimériques entre les isolats primaires du sous-type A2 du VIH-1/M et le RBF 206 du VIH-1/O ont été construites à travers une PCR qui se chevauche (appelée PCR reconstitutive) comme décrit précédemment[184,200]. En bref, chaque gène chimérique est obtenu à partir de deux fragments (fragment 5' et fragment 3' du gène IN), d'une origine phylogénétique donnée. Les amorces correspondant à la séquence interne (appelées amorces internes, sens et anti-sens) du gène, se chevauchent dans la séquence où le changement d'origine phylogénétique est souhaité et sont complémentaires, ce qui permet l'hybridation des deux fragments (fragments 5' et 3') (Figure **25.A**). Ces amorces sont utilisées conjointement avec un couple d'amorces aux extrémités de la cassette IN, appelées amorces externes (sens et anti-sens), pour l'amplification par PCR indépendante de chaque fragment (activation de la polymérase, 3min/95°C ; [dénaturation, 30sec/95°C ; hybridation, 30sec/60°C ; élongation, 1min/72°C] x30). La polymérase et le mix réactionnel utilisés proviennent du kit Phusion (New Englands Biolabs). Ces fragments (5' et 3') sont ensuite mélangés dans une PCR ultérieure où les amorces ne sont ajoutées qu'après quatre cycles (activation de la phusion, 3min/95°C ; [dénaturation, 2min/95°C ; hybridation, 2min/55°C ; élongation, 2min/72°C]*4) durant lesquelles la polymérase utilisera les régions

complémentaires hybridées comme amorce pour reconstituer le gène chimère (Figure **25.C**). Le gène chimère entier reconstitué est ensuite amplifié par PCR (activation de la polymérase, 3min/95°C ; [dénaturation, 30sec/95°C ; hybridation, 30sec/60°C ; élongation, 1min/72°C] x26) après l'ajout des amorces externes (Figure **25.D**). Les différents gènes chimériques sont digérés avec BspEI et SalI et clonés dans le plasmide p8.91. Un protocole similaire a été utilisé pour produire des intégrases mutantes : les mutations souhaitées ont été directement insérées dans la séquence des amorces internes sens et anti-sens. Le gène muté entier est obtenu et cloné comme indiqué ci-dessus.

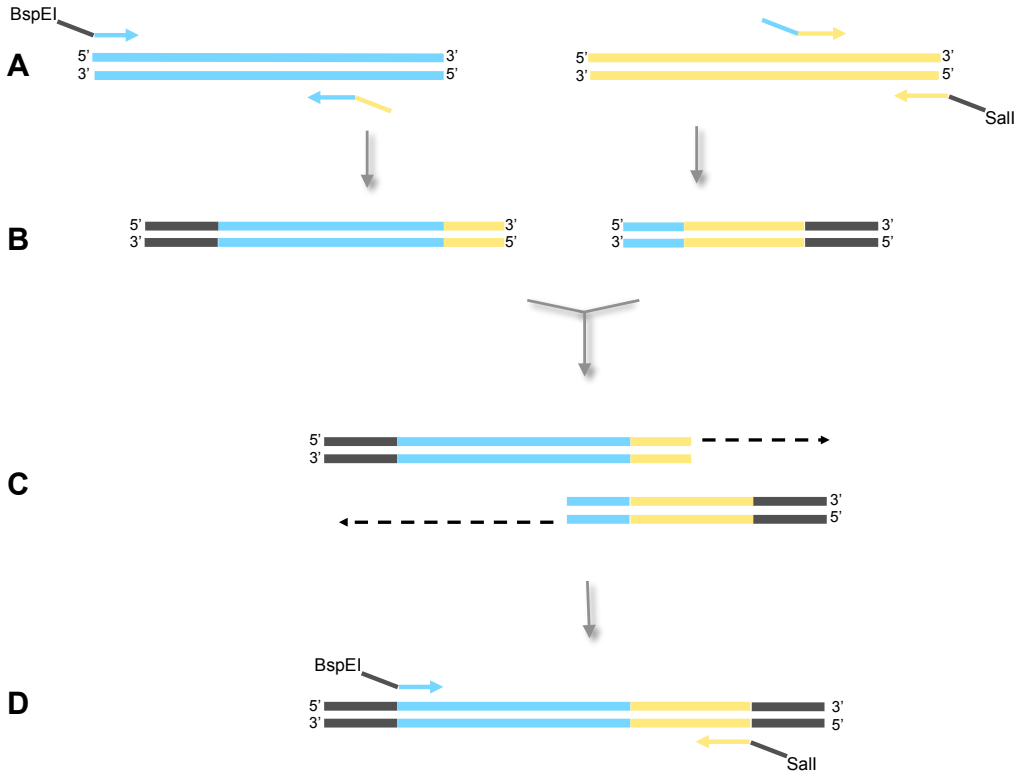


Figure 25 : Schématisation de la PCR reconstitutive. A. Les deux gènes parentaux sont indiqués en bleu et en jaune. Les amorces externes (aux extrémités du gène) sont composées en 5' des sites de restriction pour BspEI et SalI et en 3' des séquences complémentaires aux gènes parentaux. Les amorces internes (à l'intérieur du gène) sont complémentaires et s'hybrident au niveau du point de recombinaison entre les deux gènes. **B.** Les fragments 5' (à gauche) et 3' (à droite) qui permettent de reconstituer le gène chimère sont produits indépendamment par PCR. **C.** Les fragments 5' et 3' sont mélangés dans un mix réactionnel de PCR, ils s'hybrident par leurs extrémités complémentaires et subissent quelques cycles de PCR. La polymérase allonge les extrémités de chaque fragment afin de reconstituer le gène chimère. **D.** Le gène chimère est amplifié par PCR après ajout des amorces externes.

3. Cellules et souches virales

Les cellules HEK 293T ont été obtenues auprès de l'American Type Culture Collection (ATCC) et cultivées dans le milieu Dulbecco's Modified Eagle's (Gibco) additionné de 10% de sérum de veau fœtal et de 100 U/ml de pennicilline et de 100 mg/ml de streptomycine

(Thermo Fisher) à 37 ° C dans 5% de CO2.

Nous avons utilisé des isolats primaires du sous-type A2 (numéro d'accession GenBank AF286237, nommé ci-après isolat A) et du sous-type C (numéro d'accession GenBank AF286224, nommé ci-après isolat C) du VIH-1/M obtenu à partir du programme NIH AIDS Research and Reference Reagent Program. Nous avons également utilisé un CRF02_AG (AAS638), un isolat primaire du sous-type B du VIH-1/M (AiHo, appelé ci-après isolat B) et l'isolat primaire RBF 206 (Genbank accession # KU168298, nommé ci-après isolat O) du VIH-1/O, obtenu du Dr. Plantier, Unité de Virologie au CHU de Rouen associée au Centre national de référence du VIH.

4. Génération de particules virales pseudotypées

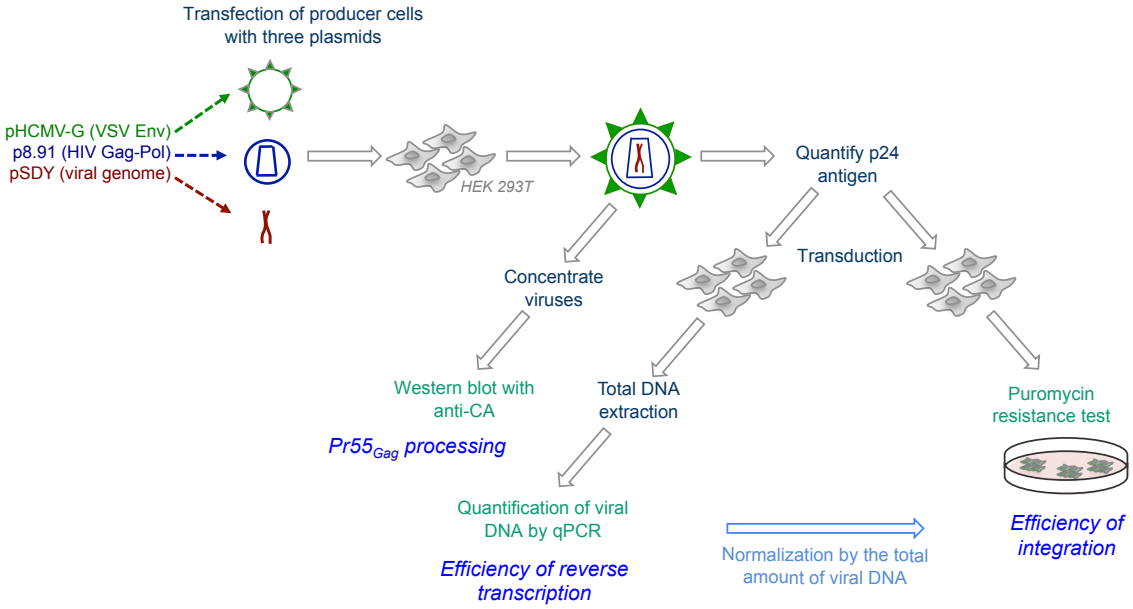


Figure 26 : Schématisation du système expérimental permettant les tests fonctionnels des virus portant les IN parentales, chimères ou mutantes.

Les particules lentivirales pseudotypées ont été produites par co-transfection de cellules HEK 293T avec le plasmide **p8.91** et deux autres plasmides de transcomplémentation, un codant pour l'ARN génomique (**pSDY**) et le plasmide **pHCMV-G** décrit dans la littérature[316] qui code pour la glycoprotéine G de l'enveloppe du virus de la stomatite vésiculeuse. La transfection est effectuée avec la méthode de la polyéthylénimine (PEI, MW 25000, linéaire, Polysciences, Warrington, PA, USA). Des cellules HEK 293T ($5x10^6$) ont été ensemencées dans des boîtes de Pétri de 100 mm de diamètre et transfectées 16-20 h plus tard. Le milieu des cellules a été remplacé par du milieu frais 6 h après la transfection. Les surnageants viraux ont été récupérés 48 à 72 h plus tard et filtrés sur filtre de 0,45 µm (Millipore). La quantité de p24 (CA) présente dans les surnageants a été quantifiée par un test ELISA

(InnotestTM HIV Antigen mAB, International Genetic Technologies, Chilly- Mazarin, France). Les virions sont ensuite utilisés pour des tests fonctionnels qui permettent, entre autres, la quantification du clivage de Pr55Gag, de l'efficacité de transcription inverse et de l'efficacité d'intégration (Figure **26**).

5. Western blot

Une analyse par Western blot a été effectuée sur une partie de la production virale pour évaluer la protéolyse des précurseurs Pr55Gag et Pr160Gag-Pol. Un volume de 1,5 ml de surnageant viral a été concentré par centrifugation (2h, 20000rcf, 4°C) sur un coussin de 20% de saccharose, et le culot de virions a été lysé dans du tampon Laemmli 1,5X. Le lysat total de virions a été migré pendant 45 minutes sur un gel gradient 4-15% de TGX (Biorad), puis transféré sur une membrane PVDF (Tampon TGS_{1x}-$Ethanol_{10\%}$, 200 mA, 2h). La membrane est bloquée sur la nuit à 4°C dans une solution de blocage (Lait 5% - PBS_{1x}-$Tween_{20\%}$), puis incubée (1h, température ambiante) avec un anticorps monoclonal de souris dirigé contre la capside (NIH AIDS Research and Reference Reagent Program, # 3537), afin de détecter la capside virale mature, la polyprotéine Pr55Gag non clivée et les intermédiaires protéolytiques contenant la CA. Un anticorps secondaire anti-souris conjugué à la HRP (Sigma-Aldrich) a été utilisé pour sonder la membrane préalablement incubée avec l'anti-CA. La membrane a ensuite été incubée avec le réactif ECL (Thermo Fisher) et la chimioluminescence a été détectée avec l'appareil Chemidoc Touch (Biorad). L'intensité des bandes a ensuite été estimée grâce au logiciel ImageLab (Biorad), afin d'exprimer le taux de CA mature par rapport à la quantité de CA totale (Pr55Gag et intermédiaires de clivage et Ca mature) (Figure **26**).

6. Transduction et évaluation des fonctions virales

Deux tests fonctionnels ont été utilisés, l'un pour surveiller l'activité de transcription inverse (dosage par qPCR), l'autre pour quantifier le taux d'ADN proviral (dosage de la résistance à la puromycine) (Figure **26**).

- Dosage de l'activité de transcription inverse par qPCR

L'ADN non internalisé a été éliminé par traitement des surnageants de virions avec 200 U/mL de Benzonase® (Sigma-Aldrich) en présence de 1 mM MgCl2 pendant 1 h à 37°C. Des cellules HEK 293T ($0,5x10^6$) ont été transduites par spinoculation (2h, 32°C/800rcf), avec un volume de virions correspondant à 200 ng de p24 et 8 pg/mL de polybrène (Sigma-Aldrich). Après centrifugation, le surnageant a été retiré et les cellules ont été remises en suspension dans un volume de 2 mL de DMEM puis étalées dans des plaques 6 puits. Après 30 h, les cellules ont été trypsinées et culotées. L'ADN total a été extrait avec le kit

UltraClean® GelSpin® DNA Extraction Kit (Ozyme), et un plasmide contenant la séquence GST en tant que contrôle interne (pGex-GST) a été ajouté au lysat. Deux tests de qPCR en duplex ont été réalisés, l'un pour la quantification et l'autre pour la normalisation. La première qPCR en duplex permet de quantifier, en parallèle, les produits précoces de la transcription inverse (strand strong stop) par détection de la jonction R-U5, et les produits tardifs (quantité totale d'ADN viral) avec la détection de la jonction U5-Ψ (Figure **27**).

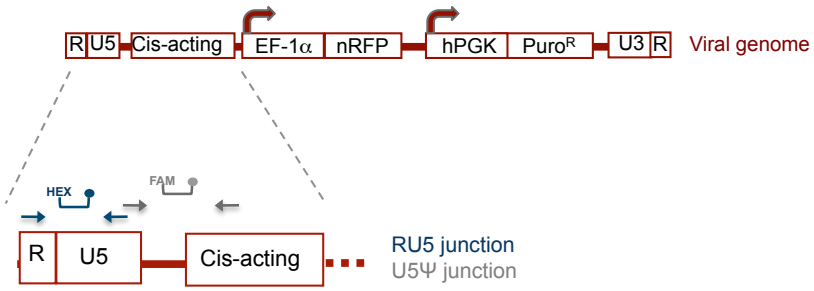


Figure 27 : Représentation de l'hybridation des amorces et sondes sur le génome viral. Les éléments du génome viral sont indiqués en marron. L'agrandissement de la partie R U5 Cis-acting montre les zones d'hybridations des amorces et sondes utilisées pour les qPCR. Les amorces (flèches) et sondes Taqman (crochet, le cercle représentant l'extincteur en 3' et HEX ou FAM, le fluorophore) sont représentées, en bleu, pour la détection de la jonction RU5 et en gris, pour la détection de la jonction U5ψ.

L'autre duplex permet de quantifier, en parallèle, la quantité de cellules (détection de l'exon 6 de la β-actine cellulaire) et les fluctuations de l'efficacité de l'extraction d'ADN (détection de la séquence GST). Les tests qPCR ont été conçus avec la technologie de sonde d'hydrolyse Taqman® à l'aide du logiciel de conception IDT Primers and Probes (International DNA technologies, Leuven, Belgium / PrimerQuest Tool), avec des sondes double extincteurs (un extincteur interne ZENTM et un extincteur en 3' Iowa BlackTM FQ). Toutes les amorces et sondes ont été synthétisées par IDT (Tableau **1**).

Les QPCR ont été réalisées avec le Supermix de sondes universelles iTaq (Biorad) sur un thermocycleur CFX96 (Biorad) avec les conditions de cycles suivantes: activation initiale de la Taq 3min/95 °C - [dénaturation 10sec/95°C, élongation 20sec/ 55°C] x40.

L'analyse des courbes des standards a été effectuée avec le logiciel CFXManager (Biorad). Les quantités tardives ou précoces d'ADN ont été normalisées avec les deux contrôles, la β-actine et la GST. Le nombre de copies d'ADN a été déterminés en référence à une courbe standard préparée par des dilutions en série des plasmides correspondant (pJet-1LTR, pGex-Gst).

duplex	target	primer/probe	sequence (5'-3')	fluorophore
duplex I (quantification)	U5Ψ	U5Ψ-F*	GTGACTCTGGTAACTAGAGA	-
	U5Ψ	U5Ψ-probe	CGCTTTCAAGTCCCTGTTCGGG	FAM
	U5Ψ	U5Ψ-R**	GAGAGCTCCTCTGGTTTC	-
	RU5	RU5-F	CAGATCTGAGCCTGGGAG	-
	RU5	RU5-probe	AAGCAGTGGGTTCCCTAGTTAGCC	HEX
	RU5	RU5-R	GGCACACACTACTTGAAGC	-
duplex II (normalisation)	GST	GST-F	CGTTATATAGCTGACAAGCACAAC	-
	GST	GST-probe	AGAGCGTGCAGAGATTTCAATGCTTG	FAM
	GST	GST-R	GCAATTCTCGAAACACCGTATC	-
	ACTB	IDT pre-designed assay, Hs.PT.56a.40703009.g/exon 6		HEX

* forward
** reverse

Tableau 1 : Sondes et amorces utilisées pour les qPCR en duplex.

- Dosage de l'ADN proviral par test de résistance à la puromycine

Des cellules HEK 293T ($0,5x10^6$) ont été transduites avec un volume de vecteur correspondant à 0,2 ng de p24 par spinoculation (2h, 32°C/800rcf), avec 8µg/mL de polybrène (Sigma-Aldrich). Après centrifugation, le surnageant a été retiré et les cellules ont été remises en suspension dans un volume de 7mL de DMEM puis étalées dans des boîtes de Pétri de 10mm de diamètre. Après 30 h, la puromycine a été ajoutée à une concentration finale de 0,6 µg/mL. Les cellules sont laissées en culture durant 10 à 12 jours puis le milieu est retiré et les clones résistants à l'antibiotique sont dénombrés. Les résultats du test de résistance à la puromycine sont ensuite normalisées par la quantité d'ADN viral tardif détectée, afin d'exprimer l'efficacité d'intégration (Figure **26**).

- Dosage de l'ADN proviral par Alu PCR

stage	target	primers/probes	sequence (5'-3')	fluorophore
1st PCR	ALU-LTR	Alu forward	TGCTGGGATTACAGGCGTGAG	-
	ALU-LTR	Ψ reverse	GCTCCTCTGGTTTCCCTTTC	-
2nd qPCR	RU5	RU5-forward	see above, table 1	
	RU5	RU5-probe		
	RU5	RU5-reverse		

Tableau 2 : Sondes et amorces utilisées pour le test d'intégration par Alu pCR.

L'ADN total extrait des cellules transduites pour le dosage de qPCR a été utilisé pour le test d'intégration par Alu PCR, comme déjà décrit dans la littérature[317]. En bref, deux cycles d'amplification ont été utilisés. Le premier cycle de PCR, avec l'amorce sens qui s'hybride sur les séquences Alu et l'amorce anti sens sur la séquence ψ de l'ADN viral, amène à l'amplification des fragments Alu-LTR (Tableau **2**). Les conditions de cyclage sont les

suivants: 95 °C pendant 3min, [95 °C pendant 30s, 55 °C pendant 30s, 72 °C pendant 3min30s] x15, 72 °C pendant 7min. Les échantillons ont été dilués à 1:10 et 2µL ont été utilisés dans la deuxième amplification par qPCR, pour détecter la jonction R-U5 des LTR, comme décrit ci-dessus pour le dosage de l'activité de la transcriptase inverse.

7. Quantification des cercles à 2LTRs par qPCR

Les cercles à 2LTRs ont été quantifiés par qPCR comme décrit précédemment[265,266]. Les virions ont été traités à la Benzonase® afin d'éliminer l'ADN non internalisé, comme pour le dosage de l'activité de transcription inverse (voir ci-dessus). Des cellules HEK 293T ($0{,}5x10^6$) ont été transduites avec un volume de vecteur correspondant à 1 µg de p24 par spinoculation puis étalées dans des plaques 6 puits comme décrit ci-dessus. Après 30 h, les cellules ont été trypsinées et culotées. L'ADN total a été extrait avec le kit UltraClean® GelSpin® DNA Extraction Kit (Ozyme). Les produits tardifs de la transcription inverse (détection de la jonction U5-Ψ) ont été évalués comme décrit en amont. Deux tests de qPCR ont été réalisés pour quantifier d'une part la quantité totale de cercles à 2LTRs, relative à l'efficacité d'import nucléaire, et d'autre part, grâce avec une amorce chevauchant la jonction des LTRs, la quantité de cercles à 2LTRs avec une jonction palindromique parfaite, relative à l'efficacité de clivage en 3' des LTRs (première étape de l'intégration). Les tests qPCR ont été conçus avec la technologie de sonde d'hydrolyse Taqman® en utilisant le logiciel de conception IDT Primers and Probes. Toutes les amorces et sondes ont été synthétisées par IDT (Tableau **3**). Les QPCR ont été réalisés avec le Supermix de sondes universelles iTaq (Biorad) sur un thermocycleur CFX96 (Biorad) avec les mêmes conditions décrites en amont. Les nombres de copies des différentes formes d'ADN viral ont été déterminées en référence à une courbe standard préparée par des dilutions en série du plasmide correspondant (pGenuine2LTR).

target	primers/probes	sequence (5'-3')	fluorophore
2LTRc	2LTR forward	CCCTTTTAGTCAGTGTGGAA	-
2LTRc	2LTR probe	TTCACTCCCAACGAAGACAAGATATCCTT	FAM
2LTRc	2LTR reverse	GTAGCCTTGTGTGTGGTAGA	-
PJ	2LTR PJ forward	TGTGGAAAAATCTCTAGCAGTAC	-
PJ	2LTR probe	TTCACTCCCAACGAAGACAAGATATCCTT	FAM
PJ	2LTR reverse	GTAGCCTTGTGTGTGGTAGA	-

Tableau 3 : Sondes et amorces utilisées pour les qPCR détectant les cercles à 2LTR**s.**

8. Tests statistiques

Toutes les analyses statistiques ont été effectuées sur au moins trois expériences indépendantes (transfection et transduction). L'analyse statistique a été effectuée avec le logiciel Prism 6 (GraphPad). Toutes les valeurs obtenues pour les chimères ou les mutants ont été normalisées en utilisant les valeurs obtenues pour le parent A, à l'exception de la quantification des 2LTRs (totaux et à jonctions parfaites), exprimés en fonction de l'IN A catalytiquement inactive ($D_{116}A$). Les différences significatives ou non par rapport au parent reflètent des test bilatéraux de Student d'écart à une moyenne théorique, afin d'évaluer si les moyennes des valeurs normalisées obtenues avec les intégrases chimères et mutantes sont significativement différente de celle obtenue avec la souche parentale.

Les quantités d'ADN précoces et tardifs sont exprimées par rapport au parent A. Un test bilatéral de Student d'écart entre deux moyennes théoriques a été effectué pour chaque échantillon entre les moyennes des quantités d'ADN précoces et tardifs afin de détecter une éventuelle différence, illustrant un défaut de transcription inverse.

9. Alignements de séquences

Près de 4000 séquences ont été utilisées pour les alignements de séquences. Les séquences originaires du VIH-1/M ont été téléchargées à partir de la base de données LANL (Los Alamos National Laboratory). Elles proviennent de différents sous-types : A (249 séquences), B (2450 séquences), C (450 séquences), D (121 séquences), G (80 séquences), H (8 séquences), J (6 séquences), K (2 séquences).

Les séquences originaires du VIH-1/O (48 séquences) proviennent d'une part de la base de données de LANL et d'autre part du centre de référence de VIH, CHU de Rouen (en collaboration avec l'équipe de Jean-Christophe Plantier).

Les séquences d'origine VIH-2 (564 séquences, tous groupes confondus) ont été téléchargées à partir de la base de données LANL.

Les alignements de séquences ont été effectués avec le logiciel CLC sequence viewer 6.

Les logos illustrant la conservation de séquences ont été effectuées avec le logiciel WebLogo 2.8.2.

Résultats et Discussions

I. Analyses des chimères intergroupes

II. Caractérisation des réseaux de coévolution au sein du NTD et du CCD

III. Le motif NKNK du CTD

I. Analyses des chimères intergroupes

L'intégrase est une des enzymes essentielles du cycle de réplication du VIH et, à ce titre, une cible de la thérapie antirétrovirale. Pourtant, beaucoup de zones d'ombres restent encore à explorer pour comprendre le fonctionnement de cette protéine. Ce travail vise à comprendre quelles sont les contraintes coévolutives qui limitent la variabilité de l'intégrase. En effet, les protéines du VIH sont soumises à une grande variabilité et doivent probablement être le siège de réseaux de coévolution, permettant d'allier la préservation de la fonctionnalité avec la diversification de leurs séquences.

La construction de chimères permet le réassortiment des polymorphismes génétiques entre deux intégrases d'origines phylogénétiques différentes. Si les réseaux de coévolution entre acide aminés ont évolué de façon spécifique au sein de chaque groupe, ils pourraient être perturbés dans les chimères et avoir un impact sur l'infectivité des virus. Afin d'exploiter les réseaux de coévolution pour étudier la fonction des différentes parties de la protéine (IN), nous avons construits des chimères entre des intégrases phylogénétiquement distantes, comme celles appartenant aux groupes M et O du VIH-1. En effet, plus la variabilité entre les séquences est grande, plus la probabilité d'observer une perturbation des réseaux de coévolution est élevée, et les isolats primaires provenant des groupes M et O partagent environ 80 % d'identité de séquences au niveau de l'IN.

Résultats

1. Construction des IN chimères A/O

Onze intégrases chimères ont été construites comme décrit dans la partie **Matériels et méthodes.** Deux isolats primaires ont été utilisés pour les construire, l'un du sous-type A du groupe M, appelé par la suite isolat A, l'autre, originaire du groupe O, est appelé isolat O. Ces chimères, représentées dans la figure **28.A**, sont construites à partir de l'IN O où des régions allant du N-ter jusqu'au C-ter de l'IN sont successivement remplacées par la séquence d'origine A afin d'augmenter progressivement le nombre de résidus qui diffèrent (cinq résidus en moyenne) entre les séquences A et O. Les chimères sont nommées par la position du résidu à partir duquel la séquence change d'origine, indiquée à droite du schéma (1 à 285, Figure **28.A**). Les gènes chimères ont ensuite été clonés dans le plasmide p8.91 parental A (contenant la RT et l'IN d'origine A), en aval de la séquence de la RT A (les séquences codantes de Gag et PR sont d'origine HXB2, celle du p8.91 initial, Figure **23**). Le parental O contient également les séquences codantes de Gag et PR d'origine HXB2 mais les cassettes RT et IN ont été remplacées par celles provenant de l'isolat O. Ces plasmides sont par la suite utilisés pour la production de particules virales.

La fonctionnalité des intégrases chimères a été évaluée par des tests de transduction comme représentée dans la figure **26**. Les vecteurs lentiviraux ont été produits par triple transfection de cellules avec le plasmide pHCMV-G[316] (codant pour la protéine d'enveloppe du VSV), le pSDY[315] (qui amène à la synthèse de l'ARN génomique viral) et le p8.91 (codant pour les polyprotéines Gag et Gag-Pol), qui contient les séquences des intégrases parentales ou chimères. Les virions sont ensuite utilisés pour transduire des cellules en culture. Comme l'ARN génomique ne contient pas les gènes viraux, l'infection sera bloquée après l'intégration de l'ADN proviral, on parle alors de système en cycle unique. L'ARN génomique contient le gène *pac*, sous contrôle du promoteur interne de la phosphoglucokinase humaine (Figure **27**), codant pour la puromycine N-acétyl transférase et permettant de conférer la résistance à la puromycine à toutes cellules infectées ayant intégrées l'ADN proviral. L'idée est que, une fois que la transcription inverse est terminée, si l'intégration se produit, l'ADN proviral conduira à la production de la puromycine N-acétyl-transférase et les cellules individuelles pourront se développer en présence de puromycine, conduisant à la génération de clones. L'expression transitoire du gène à partir de formes non intégrées de l'ADN reverse transcrit, en revanche, ne conduit pas, à long terme, à la génération de clones en présence de puromycine.
Puisque des mutations dans l'IN peuvent également affecter des étapes du cycle autres que

l'intégration, comme par exemple la transcription inverse et la maturation, le bon déroulement de ces étapes est également contrôlé en plus de l'intégration (Figure **26**).

La fiabilité de cette approche a été mise en évidence avec l'analyse de quatre variants du p8.91. Les deux parents (RT-IN des isolats A et O), appelés $RTA^{(+)}$ $INA^{(+)}$ et $RTO^{(+)}$ $INO^{(+)}$ dans le tableau **4**, et le parent A portant soit une double mutation dans la RT, la rendant catalytiquement inactive (D110N-D185N) et appelé $RTA^{(-)}$ $INA^{(+)}$ soit une mutation ponctuelle dans la séquence de l'IN, la rendant catalytiquement inactive, (D116A) appelé $RTA^{(+)}$ $INA^{(-)}$. Comme le montre le tableau **4**, tous les échantillons ont donné les résultats attendus en qPCR et dans le test de résistance à la puromycine, en fonction des propriétés de leur RT et IN respectives. En parallèle, pour corroborer davantage les résultats obtenus avec le test de résistance à la puromycine, une analyse de l'intégration par Alu PCR a été effectué[317], montrant une concordance parfaite entre les deux quantifications (tableau **4**).

	RT (QPCR, % HXB2)	**IN** ($puro^R$ clones, % HXB2)	**IN** (Alu PCR, % HXB2)
$RTA^{(+)}$ $INA^{(+)}$	90 ±21	60 ±11	69 ±14
$RTO^{(+)}$ $INO^{(+)}$	51 ±19	42 ±16	39 ±9
$RTA^{(+)}$ $INA^{(-)}$	71 ±37	0 ±0	0 ±0
$RTA^{(-)}$ $INA^{(+)}$	0 ±0	0 ±0	0 ±0

Tableau 4: Validation du système expérimental. Activité de la RT (détectée par qPCR) et de l'IN (détectée par le test de résistance à la puromycine et le test par Alu PCR) des deux isolats parentaux (A et O), et de mutants catalytiques de la RT (double mutant D110N-D185N), et de l'IN (mutant D116A) provenant de l'isolat A, appelés respectivement $RTA^{(-)}$ et $INA^{(-)}$.

2. Tests fonctionnels des IN A/O

a. Clivage des précurseurs

Les précurseurs Pr55Gag et Pr160Gag-Pol sont clivés par la protéase mature, libérée grâce à la dimérisation du Pr160Gag-Pol. L'intégrase, ou plus précisément le domaine IN du précurseur Gag-Pol, est aussi impliquée dans l'assemblage, la maturation et la morphologie des virus. En effet, il a été montré que la délétion du domaine IN du précurseur Gag-Pol entraine des défauts de maturations dans les virus, très probablement due a une absence d'activation de la protéase virale[268,282,283]. D'ailleurs, l'absence de maturation dans les virions présentant une délétion de l'IN que nous observons dans notre système (résultat non présenté, taux de clivage de 18% et 11% comparé aux parents pour la délétion de l'IN A et O respectivement) conforte cette hypothèse. D'autres part certaines mutation ou délétion dans

le C-terminal du domaine IN du précurseur Gag-Pol sont responsables du défaut d'incorporation et maturation du précurseur dans le virus bourgeonnant[268,282,283,318].

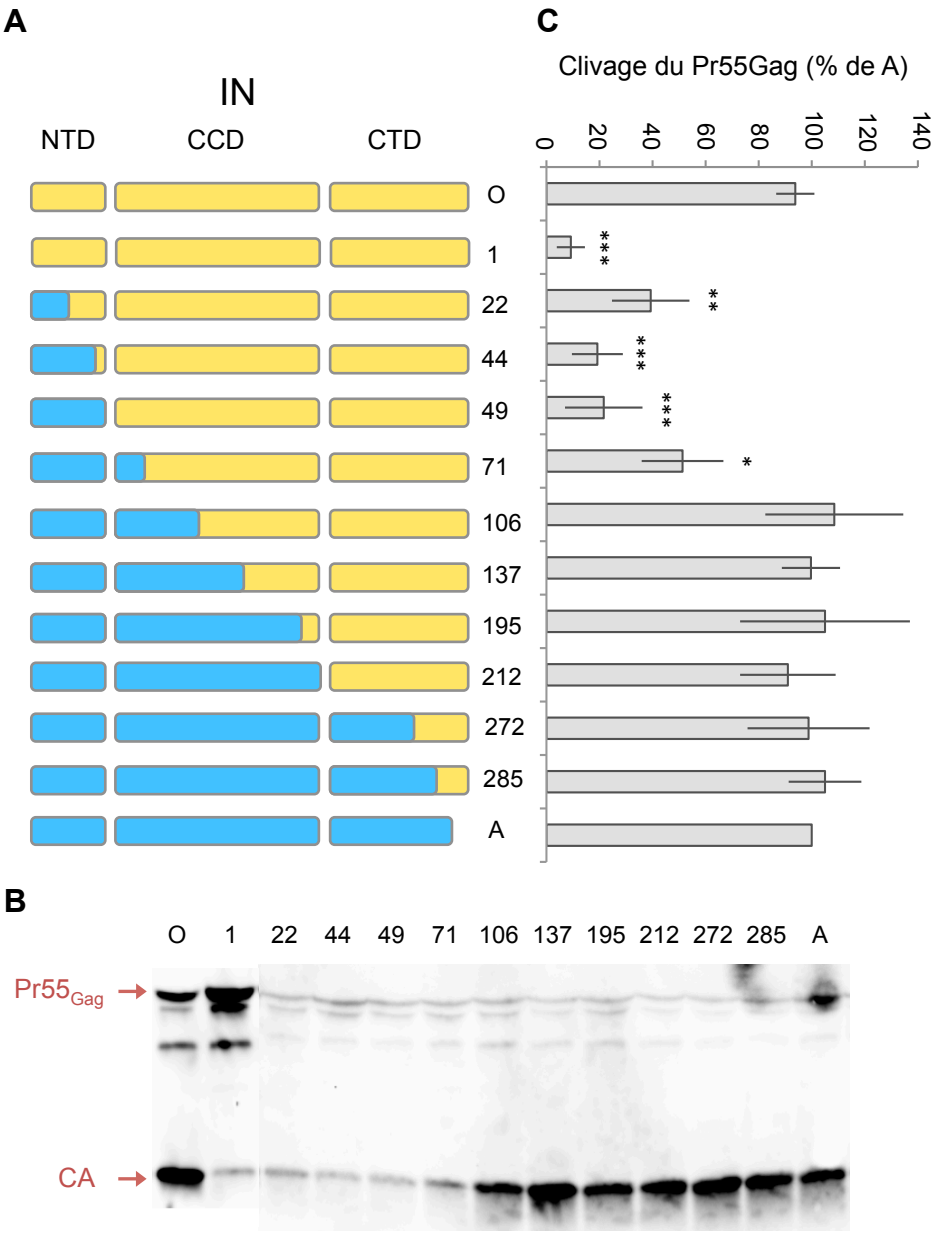


Figure 28 : Maturation des chimères AO. A. Schéma des onze intégrases chimères construites (1 à 285) ainsi que des deux parents O (en jaune) et A (en bleu). La position du point de recombinaison est indiquée à droite du schéma. **B.** Western blot avec un anticorps anti-capside. Les flèches rouges indiquent le précurseur Gag (Pr55Gag) et la capside mature (CA). **C.** Taux de clivage du précurseur Gag correspondant au rapport entre la quantité relative de la CA mature et celle du précurseur Gag et des intermédiaires de clivage, détectée par Western blot (voir B). Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé à la valeur du parent (A=100%).
(p<0,05 ; ** p<0,01 ; *** p<0,0001)*

Pour ces raisons, les éventuels effets des intégrases chimères sur la maturation sont contrôlés par quantification du taux de clivage du précurseur Pr55Gag (Figure **26**). Celui-ci est exprimé par le rapport entre la CA mature et la quantité de totale de CA (immature, au sein du Pr55Gag et de ses intermédiaires de clivage, et mature) détectée par l'anticorps anti-CA.
Le taux de clivage du Pr55Gag est comparable pour les deux parents A et O. En ce qui concerne les intégrases chimères, une diminution significative comparée au parent A (moins de 50% de l'activité de A) est observée pour les intégrases avec un point de cassure allant des positions 1 à 71 inclus. Les autres chimères (106 à 285) présentent une efficacité de clivage comparable aux deux parents (comprise entre 91% et 109%) (Figure **28.B et C**).

Ainsi, ces résultats montrent que la construction de chimères dans le NTD et le CCD entre souches d'origines phylogénétiques différentes perturbe la maturation des précurseurs polyprotéiques, probablement en interférant avec la dimérisation du précurseur Gag-Pol. De plus, certains résidus entre la position 71 et 106 semblent avoir une importance majeure dans la maturation puisque l'activité parentale est restaurée lorsque cette région est d'origine A avec la chimère 106.

b. Production de l'ADN viral

Figure 29 : Conservation de la surface d'intéraction RT-IN. Panneau gauche. Alignement de séquences réalisé avec des intégrases d'isolats du VIH - 1 (3454 séquences provenant de la base de données LANL et du centre de référence VIH, CHU de Rouen). **Panneau droit.** Alignement de séquences réalisé avec des intégrases d'isolats du VIH - 2 (564 séquences provenant de la base de données LANL). Les positions des résidus sur l'IN sont indiquées en bas (numérotation HXB2). Le logo de conservation a été effectué avec le logiciel WebLogo 2.8.2.

Comme décrit précédemment, des modifications dans la séquence peptidique de l'intégrase peuvent affecter le bon déroulement de la transcription inverse. Une interaction directe entre l'IN et la RT, faisant intervenir une surface hydrophobe localisée dans le CTD de l'intégrase serait à la base de l'implication de l'IN dans cette étape du cycle viral[272-275]. Une étude a démontré que des mutations de l'intégrase du VIH-1/M induisant une diminution de la synthèse d'ADN pouvait être complémentée par l'apport en *trans* d'une protéine de fusion IN-Vpr dans les virus mutants afin de restaurer la transcription inverse (par la RT d'origine

VIH-1/M). Cependant, puisque la complémentation avec une IN d'origine VIH-2 ne permet aucune restauration[273], il est supposé que l'IN doit être de même origine que la RT afin d'assurer cette interaction fonctionnelle.
Un alignement de séquences d'IN d'isolats primaires que nous avons effectué, montre que les résidus du CTD de l'IN constituant la surface d'interaction entre ces deux protéines sont conservés entre le VIH-1 et le VIH-2 (Figure **29**). Ainsi, l'incompatibilité entre la RT VIH-1/M et l'IN VIH-2 suggère la participation de résidus non conservés (entre VIH-1/M et VIH-2, et potentiellement entre VIH-1/M et VIH-1/O) dans le maintien de cette interaction. Le caractère chimérique des intégrases pourrait donc perturber l'efficacité de la transcription inverse.
L'impact des intégrases chimères sur la production de l'ADN viral est analysé par la quantification des produits de la transcription inverse. En plus du produit final de la transcription inverse, nous quantifions également les produits précoces, informatifs sur la nature du défaut éventuel. En effet, si la quantité de produits précoces est similaire à celle des parents alors que celle des produits tardifs est beaucoup plus basse, c'est le signe d'un défaut survenant lors ou juste après le premier transfert de brin. A l'inverse, si les produits précoces et tardifs sont en faibles quantités, le défaut d'activité concerne plus globalement l'activité de la transcriptase inverse.

Les quantités d'ADN viral détectées pour les onze chimères testées (voir Figure **28.A**) sont répertoriées dans la figure **30.A**. Comme attendu, l'efficacité de transcription inverse est quasiment indétectable (moins de 20%) pour les chimères présentant un défaut de maturation (1 à 71 inclus, Figure **28.B**).
Parmi les chimères qui ne présentent pas de défaut de clivage du Pr55Gag, certaines chimères (106 à 195) montrent une baisse significative de l'efficacité de transcription inverse (de 56% à 25%). On remarque que l'activité augmente progressivement avec le déplacement du point de cassure vers le domaine C-terminal de la protéine. Cette augmentation graduelle est corrélée avec l'augmentation de la part de séquence d'IN de même origine que la RT d'origine A et atteint une valeur similaire à celle du parent A avec la chimère 212. Ce résultat montre que le CTD d'origine O, chimère 212, est compatible avec la RT d'origine A. Par contre, lorsqu'il est chimérique (chimère 272), l'activité est affectée. En effet, la chimère 272 présente une efficacité de transcription inverse significativement basse par rapport au A (32%), alors que les deux autres chimères, 212 et 285, ont une activité quasi parentale (80% et 72% respectivement).
Ainsi, on peut supposer que des résidus non conservés présents dans les régions 106-195 et 212-285 permettent directement ou indirectement le fonctionnement de la RT, soit en régissant l'interaction RT/IN, soit en jouant un rôle sur la liaison à l'ADN dans le complexe de rétro-transcription.

La comparaison entre les quantités de produits précoces et tardifs de la transcription inverse pour chaque chimère montre qu'il n'y a pas de différence significative entre les deux types de produits (Figure **30.B**), même pour les chimères présentant un défaut de transcription inverse (voir ci dessus), à l'exception de la chimère 272. Pour cette chimère on observe une différence significative entre les quantités de produits précoces (56%) et tardifs (32%), indiquant que les évènements de transcription inverse ne vont pas à terme dû à un probable défaut dans le saut de brin ou dans la dégradation de la matrice ARN par la RNaseH, ou a une perte d'efficacité de la RT au fur et à mesure de la réaction, non détectable au niveau de la synthèse du premier fragment d'ADN de la réaction, le minus strand strong-stop DNA (-sssDNA) (voir Introduction, partie **I.3.a**). Les quantités de produits précoces et tardifs similaires mais plus bas que le taux parental pour les chimères 106 à 195 suggèrent que le défaut concerne plus globalement l'efficacité de la transcriptase inverse à convertir l'ARN génomique viral en ADN.

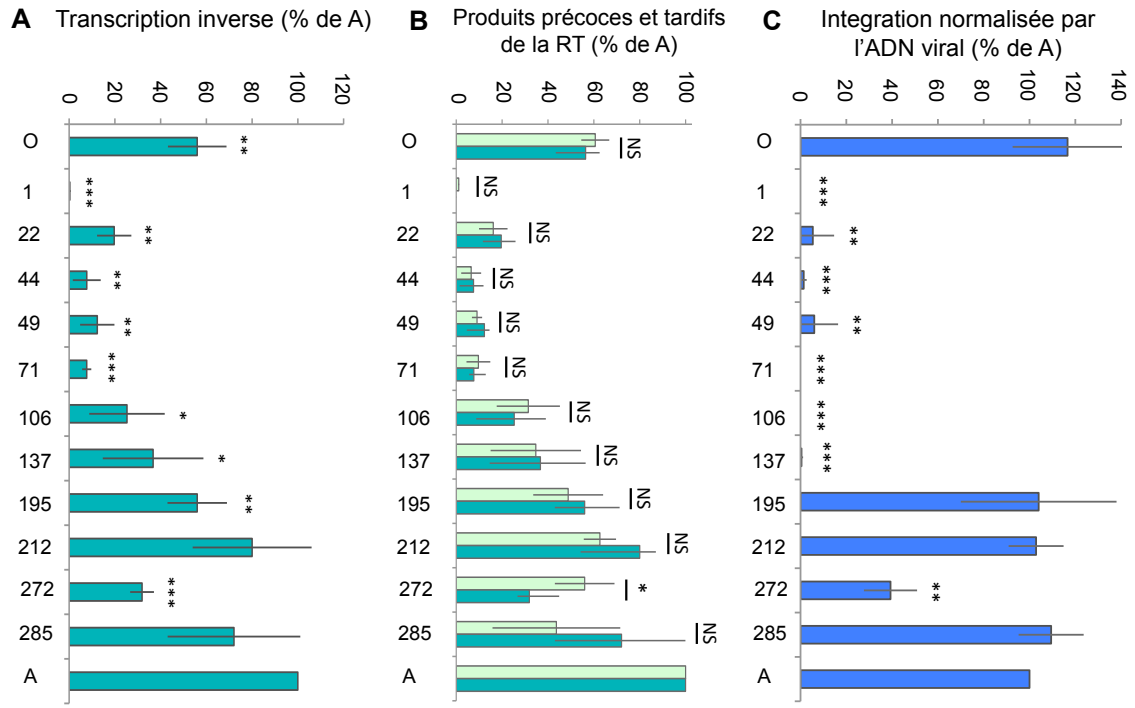


Figure 30 : Fonctionnalités des chimères AO. A. Efficacité de transcription inverse (quantité d'ADN viral tardif) comparé au parent A, fixé à 100%. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé la valeur du parent (A=100%). **B.** Comparaison entre les quantités des produits précoces (en vert clair) et tardifs (en vert foncé) de la transcription inverse, exprimées en fonction du parent A, fixé à 100%. Un test de Student à deux échantillons indique si les moyennes sont statistiquement différentes pour chaque chimère **C.** Efficacité d'intégration (nombre de clones résistants à la puromycine) normalisée par la quantité totale d'ADN viral et exprimée par rapport au parent A, fixé à 100%. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé à la valeur du parent (A=100%). Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types.
*(NS p>0,05 ; * p<0,05 ; ** p<0,01 ; *** p<0,0001)*

c. *Efficacité d'intégration*

Le caractère chimérique des IN peut évidemment affecter directement ou indirectement l'activité d'intégration. L'efficacité d'intégration est quantifiée par dénombrement des clones cellulaires résistants à la puromycine (dans lesquels l'intégration du génome viral a eu lieu de façon efficace). Comme le taux d'intégration est dépendant de la quantité d'ADN viral disponible, les données d'efficacité d'intégration sont normalisées par la quantité de produits tardifs de transcription inverse. Cette normalisation permet de mieux visualiser si les chimères/mutants de l'IN ont un impact sur la transcription inverse et l'intégration ou seulement sur la transcription inverse.

De même que pour la transcription inverse, l'activité d'intégration est pratiquement indétectable (moins de 6%) pour les chimères présentant un défaut de clivage du Pr55Gag (1 à 71 inclus) (Figure **30.C**). Les chimères 106 et 137 présentent une perte totale d'activité d'intégration qui ne peut pas être expliquée par la faible quantité d'ADN viral produit. On observe également une baisse significative de l'efficacité d'intégration pour la chimère 272 (39% de l'activité de A), alors que les autres chimères dans le domaine CTD (212 et 285) ont une activité parentale (103% et 109%, respectivement) (Figure **30.C**).

Ainsi, nos résultats montrent une incompatibilité entre les résidus d'origine O présents dans les régions 106 à 195 et 212 à 285 et le reste de l'IN A, qui, lorsqu'ils sont substitués dans par les séquences de l'isolat O, ont un impact négatif sur les efficacités de transcription inverse et d'intégration. Plusieurs régions ont été caractérisées par un défaut d'intégration ou d'une autre étape du cycle dans laquelle l'intégrase joue un rôle, cependant le défaut de maturation pour les chimères entre le NTD et le CCD n'a pas permis de conclure quant à d'éventuels effets sur la transcription inverse et l'intégration. Dans le développement qui suit, nous avons remplacé des portions de l'IN plus petites par la séquence d'origine O au sein du NTD et du CCD, afin de mieux caractériser les défauts dans le clivage des précurseurs, la transcription inverse et l'intégration.

3. Impact des IN chimères entre le NTD et le CCD sur la maturation

a. *Rôle du NTD et du CCD dans la maturation*

Nous avons tout d'abord cherché à comprendre si le caractère chimérique du site de clivage (entre RT et IN) de la protéase n'a pas d'influence sur la maturation. L'analyse de la conservation du site de clivage de la protéase entre la RT et l'IN montre qu'il est conservé chez le VIH-1/M et le VIH-1/O, à l'exception de deux résidus, l'acide aminé 558 de la RT et le 3 de l'IN (Figure **31.A**). La position 558 de la RT présente majoritairement une lysine chez le

VIH-1/M, alors que chez le VIH-1/O c'est une arginine conservée. La position 3 de l'IN porte un acide aspartique conservé chez le VIH-1/M et un acide glutamique chez le VIH-1/O. Ces différences sont retrouvées dans les séquences des deux isolats primaires A et O utilisés (Figure **31.B**).

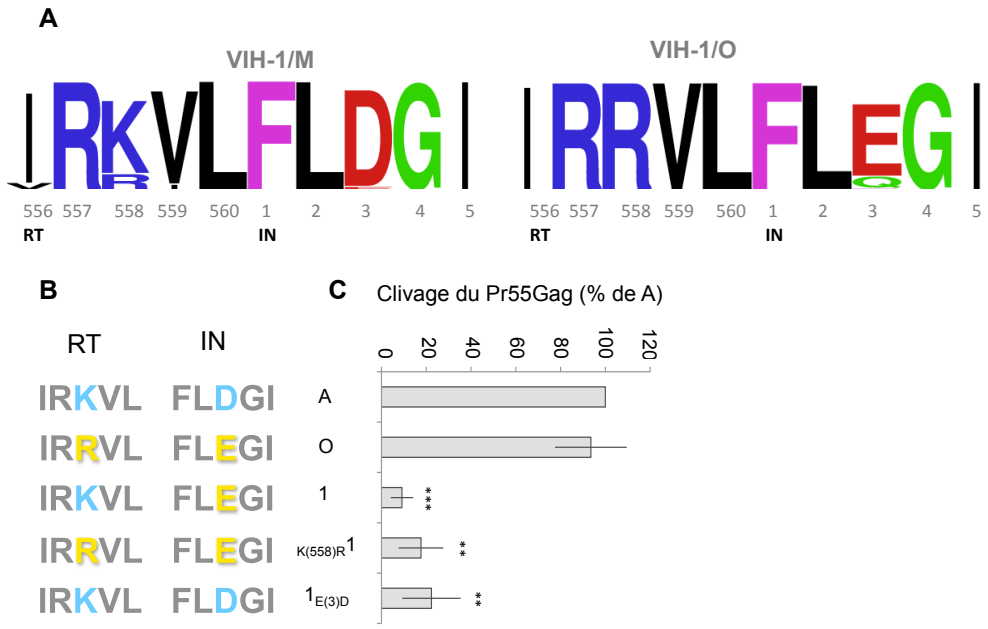


Figure 31 : Conservation du site de clivage de la PR. A. Logo de conservation de séquence du site de clivage de la protéase entre la RT et l'IN chez VIH-1/M (3366 séquences provenant de la base de données LANL) et VIH-1/O (88 séquences provenant de la base de données LANL et du centre de référence du VIH, CHU de Rouen). Les positions des résidus sur la RT et l'IN sont indiquées en bas (numérotation HXB2). Le logo de conservation a été effectué avec le logiciel WebLogo 2.8.2.
B. Représentation schématique de la séquence du site de clivage de la protéase entre la RT et l'IN. Les résidus d'origine O sont représentés en jaune, ceux d'origine A en bleu. **C.** Taux de clivage du précurseur Gag. Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé à la valeur parentale (A=100%).
(p<0,05 ; ** p<0,01 ; *** p<0,0001)*

La reconnaissance du site de clivage ayant été décrite basée sur la structure de celui-ci[138], on peut supposer que des changements de résidus peuvent perturber l'activité de la protéase. La chimère 1 présentant un faible taux de clivage de Gag par la protéase, a une lysine en position 558 de la RT (d'origine A) et un acide glutamique en position 3 de l'IN (d'origine O). Afin de tester l'implication de ces résidus dans la maturation, deux nouvelles constructions on été testées. La chimère 1 a été mutée dans le site de clivage de la protéase pour qu'il soit d'origine entièrement groupe O (chimère $_{K558R}1$) ou groupe M/A (chimère 1_{E3D}) (Figure **31.B**). Le taux de clivage de Gag de la chimère mutante 1_{E3D} n'est pas restauré par le site de clivage entièrement d'origine groupe M/A, qui devrait être parfaitement reconnu par la protéase du groupe M, ni par celui provenant du groupe O de la chimère mutante $_{K558R}1$.

Pour les deux chimères mutantes le taux de clivage est bas et similaire à la chimère 1 non mutée (Figure **31.C**).

Ces résultats suggèrent que ce n'est pas le caractère chimérique M/O du site de clivage entre la RT et l'IN de la protéase qui est responsable du défaut de maturation. On peut donc supposer que le domaine IN provenant de l'isolat O est incompatible avec le reste du précurseur Gag-Pol M (Gag et PR d'origine HXB2 et RT d'origine A), probablement due à la présence de résidus non conservés entre les deux groupes qui perturbe la dimérisation de Gag-Pol.

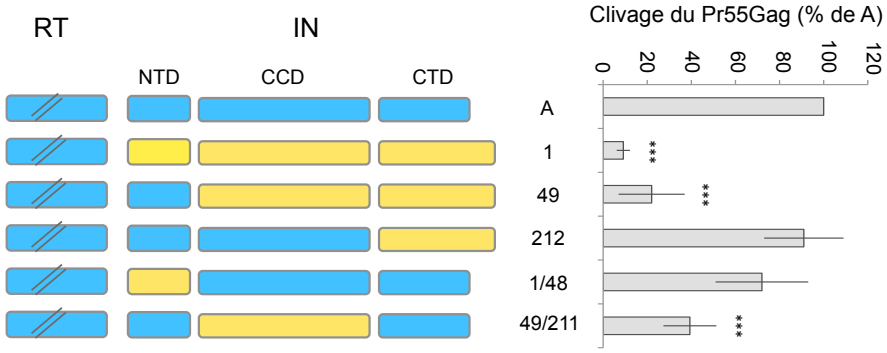


Figure 32 : Rôle du NTD et du CCD dans la maturation. Panneau gauche. Schéma des intégrases testées, clonées en aval de la RT A. Les positions du point de recombinaison ou des extrémités encadrant la région substituée par la séquence d'origine O sont indiquées à droite du schéma. **Panneau droit.** Taux de clivage du précurseur Gag correspondant au rapport entre la quantité relative de la CA mature et celle du précurseur Gag et de ses intermédiaires de clivage, détectées par Western blot. Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé à la valeur du parent (A=100%).
(p<0,05 ; ** p<0,01 ; *** p<0,0001)*

L'analyse des résultats précédents a montré que lorsque le NTD et le CCD sont de même origine que Gag-Pol (M) alors que le domaine CTD de l'IN est d'origine O, aucun défaut de maturation n'est observé (91% de l'activité parentale) (Figure **28** et **32**, 212), suggérant que les résidus non conservés entre les isolats dans ce domaine n'ont aucun rôle dans la maturation. A l'inverse, lorsque le NTD et le CCD ou seulement le CCD sont d'origine O dans le parent A, indépendamment du domaine CTD qui n'a pas d'impact sur la maturation lorsqu'il est d'origine O, le clivage des précurseurs polyprotéiques est significativement diminué comparé au taux parental (9% et 22% de l'activité de A, respectivement) (Figure **28** et **32**, 1 et 49). Ces résultats suggèrent que le NTD et/ou le CCD d'origine O sont incompatibles avec le Gag-Pol d'origine M (Gag et PR d'origine HXB2 et RT d'origine A), pour permettre la maturation efficace des précurseurs polyprotéiques.

Afin de tester l'implication du NTD et/ou du CCD dans la maturation, nous avons construit deux chimères avec uniquement le NTD ou le CCD d'origine O au sein du parent A et nous les avons nommées par les extrémités (résidu en N-ter / résidu en C-ter) de la région de l'IN

A substituée par la séquence O correspondante (Figure **32,** 1/48 et 49/211). Lorsque seul le CCD est d'origine O, le taux de clivage du Pr55Gag est bas et significativement différent de A (39% de l'activité de A), alors que lorsque seul le NTD provient de l'isolat O, l'activité est similaire au parent. Ainsi, on peut dire que le domaine CCD provenant de l'IN O est incompatible avec le précurseur Gag-Pol M (Gag et PR d'origine HXB2 et RT d'origine A), suggérant que des résidus non conservés entre les deux groupes au sein de ce domaine perturbent la dimérisation fonctionnelle de Gag-Pol nécessaire à l'autoactivation de la protéase virale.

b. Caractérisation de la coévolution au sein du NTD et du CCD

Afin de mieux définir quelle(s) région(s) est(sont) impliquée(s) dans le défaut de maturation observé avec les chimères dans les domaines NTD et CCD que je viens de décrire, sept nouvelles chimères ont été construites le long de ces deux domaines (résidus 1 à 212) avec seulement une portion du domaine de séquence d'origine O. Celles-ci sont nommées par les extrémités (résidu en N-ter / résidu en C-ter) de la région de l'IN A substituée par la séquence O correspondante (Figure **33.A**).

L'analyse du clivage des précurseurs chez les virus portant ces intégrases chimères a mis en évidence différents phénotypes (Figure **33.B**). Les chimères dans le domaine NTD (1/21 et 22/43) présentent un taux de clivage parental, tout comme la chimère dans une partie de la région flexible reliant le CCD au CTD, 196/212. A l'inverse lorsque le CCD porte des portions de séquence d'origine différente que le reste de la protéine (chimères 44/71, 72/106, 107/137 et 138/195), le taux de clivage du précurseur Gag est significativement plus bas que celui du parent A (71%, 65%, 53% et 83% respectivement) et les chimères 72/106 et 107/137 sont celles pour lesquelles la diminution est la plus marquée.

Ces résultats sont en accord avec les résultats précédents, qui montrent que l'origine phylogénétique du domaine CCD est déterminante pour la maturation, et suggèrent que les acides aminés différents entre A et O dans la région située entre les positions 72 et 106 pourraient jouer un rôle principal dans la dimérisation correcte du précurseur Gag-Pol.

4. Impact des IN chimères sur la transcription inverse et l'intégration

Etant donné que les chimères dans les domaines NTD et CCD (de la chimère 1/21 à la 196/212) présentent un taux de clivage de Gag d'au moins 50%, j'ai pu procéder à l'analyse plus fine de l'impact de chaque substitution par la région d'origine O, sur la transcription inverse et l'intégration.

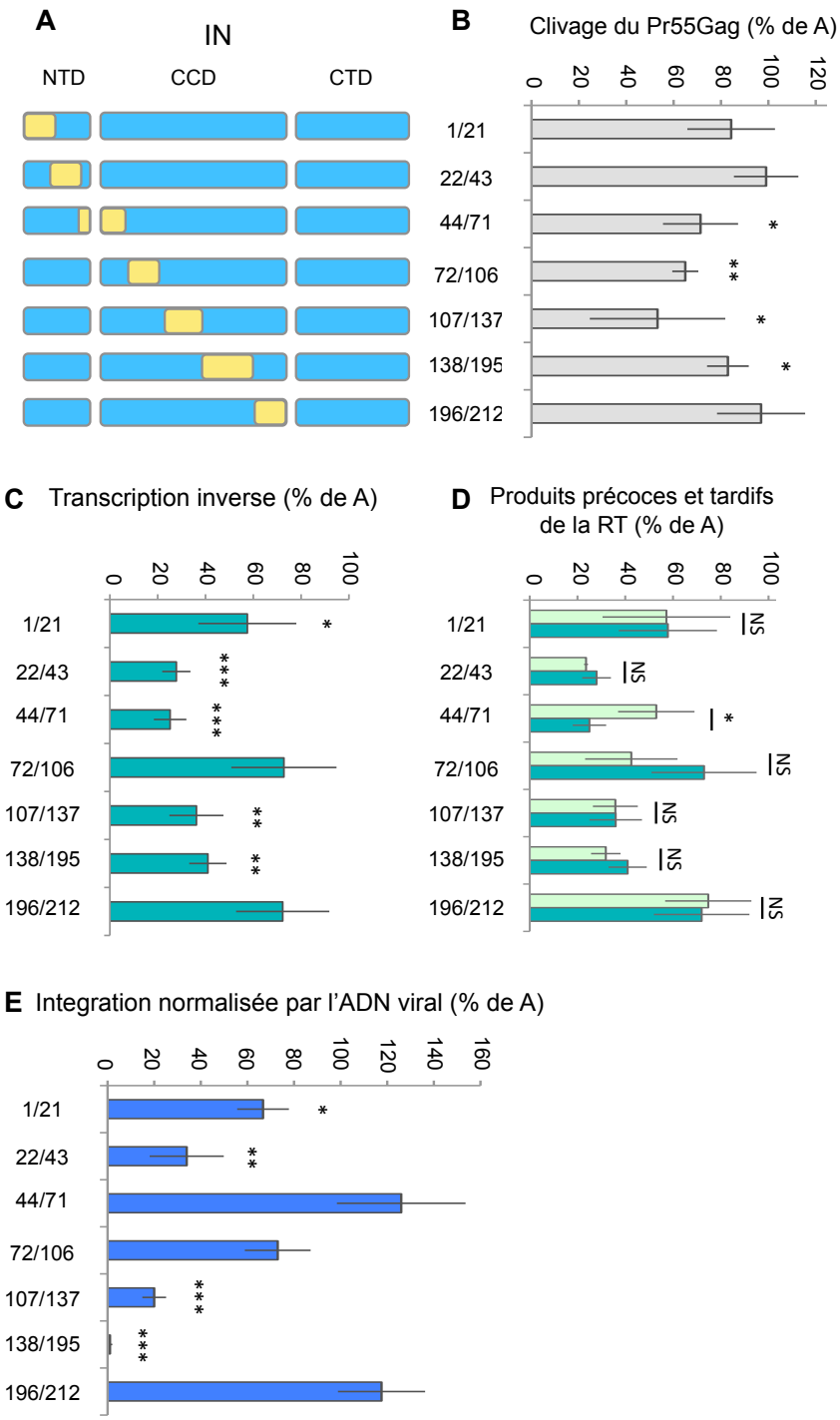


Figure 33 : Fonctionnalités des chimères AO. A. Schéma des intégrases testées, clonées en aval de la RT A. Les positions des extrémités encadrant la région substituée par la séquence d'origine O sont indiquées à droite du schéma. **B.** Taux de clivage du précurseur Gag détectée par Western blot. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé à la valeur parentale (A=100%). **C.** Efficacité de transcription inverse (quantité d'ADN viral tardif) comparé au parent A, fixé à 100%. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé à la valeur de A. **D.** Comparaison entre les quantités des produits précoces (en vert clair) et tardifs (en vert foncé) de la transcription inverse, exprimées en fonction du parent A, fixé à 100%. Un test de Student à deux échantillons indique si les moyennes sont statistiquement différentes pour chaque chimère **E.** Efficacité d'intégration normalisée par la quantité totale d'ADN viral et exprimée par rapport au parent A, fixé à 100%. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé au parental A. Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types.
*(NS p>0,05 ; * p<0,05 ; ** p<0,01 ; *** p<0,0001)*

a. *Effets sur la production de l'ADN viral*

La quantification de la production d'ADN viral des virus portant les intégrases chimériques a montré une diminution globale de l'efficacité de transcription inverse pour chaque chimère, à l'exception de celles observées avec les chimères 72/106 et 196/212 avec une activité similaire au parent (Figure **33.C**). Les cinq autres chimères testées (1/21, 22/43, 44/71, 107/137 et 138/195) présentent une baisse significative de la quantité de produit final de la transcription inverse. La taux de transcription inverse est d'environ 30% de la quantité d'ADN viral parental pour les chimères 22/43, 107/137 et 138/195, et de 22% pour la chimère 44/71. A l'inverse, la chimère 1/21 présente une baisse modérée (58% du parental A). Le remplacement de portions de l'IN A au sein du NTD et du CCD par les séquences correspondantes de l'IN O perturbe donc le bon déroulement de la transcription inverse.
Les résultats obtenus avec les chimères 22/43, 44/71, 107/137 et 138/195 montrent que des résidus au sein de ces quatre régions, non conservés entre A et O, influent positivement sur la transcription inverse, probablement par le biais de l'interaction entre la transcriptase inverse et l'intégrase.

La comparaison entre les quantités de produits précoces et tardifs de la transcription inverse pour chacune des chimères montre qu'elles sont comparables (Figure **33.D**), à l'exception de la chimère 44/71 avec une différence significative entre la quantité de produits précoces (43%) et tardifs (22%) de la transcription inverse. Cette baisse indique que près de la moitié des évènements de transcription inverse ne vont pas à terme, suggérant soit un probable défaut dans le saut de brin ou dans la dégradation de la matrice ARN par la RNaseH, soit une perte d'efficacité de la RT au fur et à mesure de la réaction, non détectable au niveau de la synthèse premier fragment d'ADN de la réaction, le minus strand strong-stop DNA (-sssDNA). Les quantités de produits précoces et tardifs similaires mais plus bas que le taux parental (1/21, 22/43, 107/137 et 138/195) indiquent un défaut qui concerne plus globalement l'activité de la RT, survenant avant la synthèse du -sssDNA.

b. *Effets sur l'intégration de l'ADN proviral*

La baisse de la production de l'ADN viral a un effet direct sur l'efficacité d'intégration, cependant la normalisation par la quantité d'ADN disponible pour l'intégration montre que le caractère chimérique de l'enzyme a également un impact sur l'activité d'intégration elle même. Parmi les sept intégrases testées, seuls les chimères 44/71, 72/106 et 196/212 ont une activité parentale (126%, 73% et 118% respectivement). Les intégrases avec des portions d'origine O dans les régions 22/43, 107/137 et 138/195 présentent un défaut d'intégration sévère et significatif, respectivement 34%, 19% et 1% de l'activité parentale (Figure **33.E**). Des résidus non conservés entre A et O dans ces régions affectent donc la

fonctionnalité de l'intégrase. Bien que plus modérée, la baisse d'efficacité d'intégration observée avec la chimère 1/21 (67%) reste néanmoins significative.

Ainsi, le test des chimères possédant au moins la moitié du taux de clivage du Pr55Gag du parent A a permis de mettre en évidence des régions de l'intégrase où se trouvent des résidus non conservés entre les isolats A et O, impliqués dans le bon déroulement de la transcription inverse et de l'intégration.

Discussion

Par la construction de chimères le long de l'intégrase entre les isolats A et O, nous avons pu mettre en évidence la présence des réseaux de coévolution jouant différents rôles dans le cycle infectieux.

1. Impact des intégrases chimères sur la maturation

Nous avons observé une baisse du taux de clivage du précurseur Pr55Gag avec certaines chimères (entre le NTD et le CCD, résidus 1 à 106) suggérant que la présence de certains résidus non conservés entre A et O (22 résidus non conservés au total dans la région 1-106) permettent, d'une façon ″groupe spécifique″, le maintien du clivage des précurseurs. Le CTD, bien qu'impliqué dans l'assemblage[285] et la maturation des virus[280,318], ne semble pas contenir de résidus non conservés important pour cette étape puisque, lorsqu'il est substitué par la séquence d'origine O dans un Gag-Pol M (chimère 212), le clivage des précurseurs polyprotéiques est au niveau du parent. En revanche, certains résidus entre les positions 71 et 106 semblent avoir une importance majeure dans la maturation puisque l'activité parentale est restaurée lorsque cette région provient de l'isolat A avec la chimère 106. D'ailleurs, les résultats obtenus avec les chimères ayant seulement le NTD (chimère 1/48) ou le CCD (chimère 49/211) d'origine O ont aussi permis d'attribuer au CCD un rôle majeur dans la maturation des précurseurs polyprotéiques. En effet, la substitution du domaine cœur catalytique de l'IN A par la séquence correspondante d'origine O perturbe le fonctionnement de la protéase, probablement via un défaut d'autoactivation due à une mauvaise dimérisation du Pr160Gag-Pol, ce qui résulte en une diminution de près de 60% du clivage du précurseur Gag. Lorsque le domaine cœur catalytique porte des portions de séquences d'origines différentes que le reste de la protéine (chimères 44/71, 72/106, 107/137 et 138/195), le taux de clivage du précurseur Gag est affecté mais l'effet des substitutions est plus modéré que celui de la substitution du domaine entier. Dans ce contexte les chimères ayant les régions 72-106 ou 107-137 d'origine O, génèrent un clivage du Pr55Gag diminué de près de moitié comparé au parent.

La région 72-106 présente 7 résidus non conservés entre les deux isolats et la région 107-137, 8. Les acides aminés de ces deux régions constituent des éléments structuraux qui se retrouvent à proximité les uns des autres dans les cristaux du CCD, qu'il soit isolé (1EXQ, IN_{52-210}, 1ITG, IN_{52-210}), lié au NTD (1K6Y, IN_{1-212}) ou lié au CTD (1EX4, IN_{49-288}), et proches de la surface de dimérisation du CCD (résidus $W_{61}E_{85}E_{87}K_{103}R_{107}W_{108}$, conservés chez les

Lentivirus)[223]. La proximité des résidus non conservés avec la région de dimérisation, lorsque ces régions sont d'origine O, pourrait résulter en une structure, adoptée par l'IN chimère au sein du précurseur Gag-Pol, qui n'est pas compatible avec la dimérisation fonctionnelle assurant l'autoclivage de la protéase, soulevant la question d'une éventuelle coévolution entre la protéase et l'IN.

Ainsi, on peut émettre l'hypothèse que la dimérisation du Pr160Gag-Pol a lieu grâce à la surface de dimérisation située au sein du CCD et que des résidus non conservés au sein de différentes régions de ce domaine (dont ceux des régions 72-106 et 107-137 seraient les plus importants) sont responsables d'un repliement caractéristique aux intégrases de chaque groupe (M et O, dans notre cas) qui aurait coévolué avec le domaine protéase du même précurseur pour assurer une dimérisation et un autoclivage optimales. D'ailleurs, le défaut sévère de maturation des chimères 1 à 106 (1 à 106, Figure 27), avec des régions d'origine O bien plus grandes, corrobore cette hypothèse, dans la mesure où plus l'IN chimère présente des résidus d'origine O, plus son repliement pourrait se rapprocher de celui adopté par l'IN O. Nous avons également pu montrer que le caractère chimérique A/O du site de clivage de la protéase entre RT et IN n'était pas impliqué dans le défaut de maturation, suggérant que le défaut de clivage du précurseur Gag de la chimère 1 (RT d'origine A et IN d'origine O) serait dû à l'incompatibilité du domaine IN d'origine O avec le reste du précurseur Gag-Pol M (Gag et PR d'origine HXB2 et RT d'origine A). Néanmoins, l'activité parentale de l'isolat O suggère que lorsque les domaines RT et IN de Gag-Pol sont tous deux d'origine O au sein du Gag-Pol M (Gag et PR d'origine HXB2), le repliement de l'IN O pourrait être compensé par celui de la RT, favorisant la dimérisation et l'autoclivage de la protéase.

Il serait intéressant de tester individuellement les résidus qui diffèrent entre les isolats A et O présents dans la région 72-106, pour leur implication dans la maturation. En effet, dans le cadre de cette étude, seules les régions présentant des défauts sévères de transcription inverse ou d'intégration ont été analysées finement et la chimère 72/106, présentant une transcription inverse et un intégration de type parentale, a donc été exclue (Figure **33.C** et **E**).

2. Effets des intégrases chimères sur la transcription inverse

Nous avons également pu constater une diminution de la quantité d'ADN viral produit par transcription inverse pour certains virus portant les IN chimères. Bien évidemment, pour les chimères présentant un très fort défaut de maturation, il n'est pas possible de conclure quant à un éventuel effet du caractère chimérique de l'intégrase sur la transcription inverse, due à la probable absence d'IN et de RT matures dans les virions. Cependant, certaines chimères

avec un taux de clivage des précurseurs polyprotéiques comparables aux virions parentaux présentent un défaut de transcription inverse. Tout d'abord, nous avons mis en évidence, par l'analyse des chimères intergroupes, que des résidus non conservés entre A et O présents entre les positions 106 et 212 sont impliqués dans le maintien de la transcription inverse, probablement par le biais de l'interaction RT-IN. De plus, les résultats obtenus avec les chimères qui portent des portions de séquences d'environ 30 résidus d'origines différentes du reste de la protéine, présentant des taux de maturation d'au moins 50%, ont permis de mieux caractériser les régions de l'IN (non maturés dans les chimères avec des portions plus grandes d'origine O) entre le NTD et le CCD (22-43, 44-71, 107-137 et 138-195) portant des résidus non conservés impliqués dans le déroulement de la transcription inverse. Ces régions, lorsqu'elles sont substituées par la séquence d'origine O, à l'exception de la région 44-71, présentent un défaut de transcription inverse qui surviendrait avant la synthèse du premier fragment d'ADN de la réaction, le minus strand strong-stop DNA (-sssDNA) (voir Introduction, partie **I.3.a**).

L'interaction entre l'IN et la RT (faisant intervenir des résidus conservés du CTD, $R_{231}W_{243}G_{247}A_{248}V_{250}I_{251}K_{258}$)[273,274] est nécessaire au bon déroulement de la transcription inverse et assure la stabilisation et le placement correct de l'amorce $tRNA^{Lys,3}$ sur la matrice[275]. Ainsi, on peut supposer que des résidus non conservés entre les IN A et O situés dans les régions de 22 à 43 et de 107 à 195, pourraient être impliqués dans le maintien de cette interaction permettant l'initiation de la transcription inverse. Comme la surface d'interaction concerne des résidus conservés du CTD, on suppose que ces résidus entre les domaines NTD et CCD affecteraient indirectement l'interaction RT-IN probablement par la structuration différente de la portion d'origine O au sein de l'IN A. Ainsi, le repliement de l'IN au sein du RTC (et donc aussi au sein du PIC) pourrait être perturbé et gêner l'interaction avec la RT qui permet d'assurer le bon déroulement de la transcription inverse.

On ne peut cependant pas exclure un problème survenant avant le processus de transcription inverse, comme un défaut dans la décapsidation ou dans la morphogénèse de la particule. En effet, les modifications dans l'IN peuvent affecter la décapsidation, par le biais du recrutement de la cyclophiline A (CypA). Comme mentionné dans l'introduction, partie **II.3.a**, la délétion de l'IN a un impact sur l'incorporation de la CypA qui, par son interaction avec la capside, défavorise la décapsidation précoce[270]. Le mauvais déroulement de cette étape peut entrainer une baisse d'efficacité de la transcription inverse, due à la dégradation du matériel génétique viral par les enzymes cellulaires, accompagnée d'un défaut dans l'import nucléaire, puisque la CypA joue également un rôle dans cette étape[108,109], résultant en une baisse plus importante de l'efficacité d'intégration. D'autres part,

des modifications dans la séquence peptidique de l'IN pourraient induire des défauts d'incorporation du complexe ribonucléoprotéique (ARN complexé à la NC) dans la particule, inhibant de ce fait l'efficacité de transcription inverse, puisque l'IN a montré être impliquée dans le recrutement du matériel génétique au sein de la capside virale via son interaction avec l'ARN génomique lors de la morphogénèse de la particule[285].

Concernant la région 44-71, nos résultats indiquent que le défaut surviendrait après le début de la synthèse d'ADN. Ainsi, ce défaut ne concerne probablement pas l'action positive qu'opère l'IN sur la RT, ni son implication dans la décapsidation et la morphogénèse, laissant supposer que ce défaut pourrait concerner une étape intermédiaire de la transcription inverse, comme par exemple la dégradation de la matrice ARN par la RNaseH, ou une perte d'efficacité de la RT en cours de synthèse. Dans ce cas, on peut supposer que la modification de la conformation de l'IN pourrait gêner le bon déroulement de la synthèse d'ADN viral. L'IN a montré se lier à deux régions de la RT, les sous-domaines doigts et paume, et la partie C-ter du sous-domaine connexion (voir introduction, partie **II.3.a**)[272]. Comme la RT et l'IN sont en interaction dès l'initiation de la transcription inverse, la liaison de l'IN pourrait perturber le placement de l'enzyme sur la matrice en cours de synthèse, si le sillon formé par les sous domaines doigts, paume et pouce (voir introduction, partie **I.2.e**)[53,54] est moins accessible dû à la liaison par l'IN dans une structure aberrante. Cet encombrement provoqué pourrait également affecter la RNaseH puisque l'IN se lie également au niveau du domaine connexion, rendant la matrice inaccessible pour le clivage de l'ARN, ce qui inhiberait le saut de brin.

D'autres régions de l'IN semblent également jouer un rôle important pour la transcription inverse. En effet, lorsque le CTD est chimérique (chimère 272), une diminution marquée de la production d'ADN viral est observée alors que lorsqu'il est entièrement d'origine O ou A, l'activité de la RT n'est pas affectée (chimère 212 et 285). Cela signifie que la portion 273-285 de l'IN O est incompatible avec la portion 212-272 de l'IN A, pour assurer la transcription inverse, puisque lorsque celles-ci sont de même origine (O avec la chimère 212 et A avec la chimère 285) l'activité est parentale. Nos résultats indiquent que le défaut surviendrait également après le début de la synthèse d'ADN, suggérant donc qu'il pourrait être dû, comme pour la région 44-71, au repliement incorrect de l'IN, liée à la RT, et gêner la transcriptase inverse dans son rôle. Il se pourrait donc que des résidus non conservés au sein des régions 44-71 et 212-285 soient impliqués dans des interactions qui permettent le maintien du repliement de l'IN, favorable à l'action de la transcriptase inverse lorsque celles-ci sont liées.

3. Impacts des chimères sur l'intégration

Enfin, en plus des défauts de maturation et de transcription inverse, nous avons également observé de sévères impacts sur l'efficacité d'intégration, suggérant que les réseaux de coévolution sont aussi essentiels pour le maintien de l'activité de l'IN. La baisse de la production de l'ADN viral a un effet direct sur l'efficacité d'intégration, cependant la normalisation par la quantité d'ADN disponible pour l'intégration permet de montrer quelles substitutions par les acides aminés d'origine O dans l'IN A ont un impact direct sur l'intégration de l'ADN viral dans le génome cellulaire.

A nouveau, pour les chimères présentant un très fort défaut de maturation, il n'est pas possible de conclure quant à un éventuel effet du caractère chimérique de l'intégrase sur l'intégration en elle-même. Cependant, certaines chimères avec un taux de clivage des précurseurs polyprotéiques comparables aux virions parentaux présentent un défaut d'intégration. Nous avons pu mettre en évidence la présence de résidus non conservés entre les isolats A et O au sein des régions de 106 à 195 et de 212 à 285 de l'IN A, impliqués dans le maintien de l'efficacité d'intégration. Les résultats obtenus avec les chimères qui portent des portions de séquences d'environ 30 résidus d'origines différentes du reste de la protéine, présentant des taux de maturation d'au moins 50%, ont permis de définir des régions (22-43, 107-137 et 138-195) entre les domaine NTD et CCD de l'IN (non maturés dans les chimères avec des portions plus grandes d'origine O) qui affectent sévèrement l'intégration. Des résidus non conservés entre les isolats A et O dans ces régions doivent probablement affecter la fonctionnalité de l'intégrase. D'ailleurs, ces résultats sont en accord avec l'observation précédente suggérant que des résidus présents dans la région de 106 à 195 interviendrait dans l'activité de l'IN. Le défaut pourrait être dû à l'implication d'acides aminés non conservés dans des interactions inter domaines, qui permettent la correcte multimérisation de la protéine, ou dans la liaison aux cofacteurs cellulaires. Au vu du rôle de l'IN dans l'import nucléaire[261-263] et de son interaction avec des protéines cellulaires permettant le passage du PIC dans le noyau (Importines 3 et 7, Transportine 2)[276-281], comme mentionné dans l'introduction partie **II.3.a**, il n'est pas exclu que ces chimères également présentent un défaut d'import nucléaire. D'ailleurs, comme plusieurs chimères présentent un défaut, à la fois dans la transcription inverse et dans l'intégration il est possible que la décapsidation soit défectueuse, puisque comme décrit précédemment, l'IN a un impact sur cette étape.

La chimère 22-43 présente uniquement un défaut d'intégration, or comme la région 22-43 porte une partie du domaine en doigt de zinc du NTD ($H_{12}H_{14}C_{40}C_{43}$), il est possible que les résidus non conservés au sein de cette région perturbent la structure du NTD, ne permettant pas le repliement correct, compatible avec la formation du domaine en doigt de zinc,

affectant ainsi l'intégration. Les chimères 107-137 et 138-195 semblent être défectueuses à la fois dans la maturation, la transcription inverse et l'intégration, puisqu'il est peu probable que la baisse modérée du clivage des précurseurs et de l'efficacité de transcription inverse soient responsables du défaut considérable d'efficacité d'intégration. Comme mentionné précédemment, il est possible que ces chimères affectent la décapsidation, résultant en une diminution à la fois dans la transcription inverse et de l'intégration. Enfin, ces deux régions étant à proximité directe de la surface d'interaction avec le cofacteur LEDGF/p75 ($A_{128}A_{129}W_{131}W_{132}I_{161}V_{165}R_{166}E_{170}L_{172}K_{173}$)[219,220] , il est possible que la présence des résidus non conservés entre A et O perturbent le repliement de l'IN, rendant inaccessible la surface d'interaction à LEDGF/p75, affectant ainsi l'intégration.

Ainsi, le test des chimères A/O a permis de mettre en évidence des régions de l'intégrase où se trouvent des résidus non conservés entre les isolats A et O, impliqués dans le maintien de l'activité catalytique de la protéine mais également dans ses rôles non catalytiques dans les autres étapes du cycle réplicatif.

II. Caractérisation des réseaux de coévolution au sein du NTD et du CCD

L'étude de l'infectivité de virus portant des intégrases chimères entre des isolats primaires de VIH-1/M et VIH-1/O que j'ai décrite dans la partie **I** a permis de mettre en évidence de potentiels réseaux de coévolution impliqués dans les activités catalytique et non catalytique de l'intégrase. Plusieurs régions ont été caractérisées par un défaut d'intégration ou d'une autre étape du cycle dans laquelle l'intégrase joue un rôle. Dans le développement qui suit, nous avons identifié les résidus présents dans le NTD et le CCD, responsables de la baisse de fonctionnalité des chimères, qui présentent des défauts à la fois dans le clivage des précurseurs, la transcription inverse et l'intégration.

Résultats

1. Identification des résidus induisant des défauts de fonctionnalité

Afin d'identifier les résidus, au sein des domaines N-terminal et cœur catalytique, impliqués dans le maintien de l'interaction fonctionnelle avec la transcriptase inverse et de la fonctionnalité de l'intégrase, nous avons testé individuellement chaque résidus qui diffère entre l'IN A et l'IN O, au sein des régions d'intérêts.

Pour l'analyse de l'efficacité de transcription inverse et d'intégration, seules les régions provoquant les diminutions d'efficacité les plus drastiques, pour au moins l'une des deux activités, lorsqu'elles sont d'origine O ont été testées : 22-43, 44-71, 107-137 et 138-195.

Les résidus de ces régions qui diffèrent entre l'isolat A et l'isolat O et leur conservation au sein de leur groupe respectif sont reportés dans le tableau **5**. Plusieurs de ces positions présentent un résidu conservé au sein du groupe M et un autre acide aminé conservé au sein du group O. Ces résidus sont des candidats idéaux pour l'appartenance à un réseau de coévolution car ils sont conservés au sein des deux groupes phylogénétiques, signe qu'ils sont probablement soumis à des contraintes fonctionnelles, spécifiques à chaque groupe.

Sous-type A groupe M	Isolat A	Position du résidu	Isolat O	Groupe O
		22-43		
M 100%	M	22	L	**L** 100%
S 97%	H	24	S	**S** 100%
N 99%	N	27	G	**G** 98%
V 99%	V	37	I	**I** 100%
D 92%	D	41	P	**P** 98%
		44-71		
Q 44%	Q	44	H	**H** 98%
L 98%	L	45	I	**I** 92%
K 100%	K	46	R	**K** 78%
G 99%	G	59	E	**E** 92%
I 71%	I	60	V	**V** 76%
L 98%	L	63	M	**M** 63%
		107-137		
V 94%	V	112	I	**I** 55%
T 97%	I	122	T	**T** 100%
A 62%	A	124	T	**A** 88%
A 99%	T	125	A	**T** 61%
V 71%	V	126	M	**M** 98%
A 100%	A	133	T	**T** 76%
Q 91%	Q	136	T	**Q** 53%
Q 99%	Q	137	H	**H** 92%
		138-195		
S 100%	S	153	A	**A** 96%
K 97%	K	160	S	**S** 100%
G 97%	G	163	Q	**Q** 91%
I 100%	I	182	V	**V** 94%
Y 99%	Y	194	C	**Y** 90%
S 95%	S	195	T	**T** 100%

Tableau 5: Conservation des résidus entre les IN A et O. Colonnes de gauche. Alignement de séquences réalisé avec des intégrases d'isolats du sous-type A du VIH-1/M (249 séquences provenant de la base de données LANL). La séquence consensus est indiquée en noir, le pourcentage indique le niveau de conservation du consensus au sein du groupe M. La séquence de l'isolat primaire A utilisé pour les tests est indiquée en bleu, dans la colonne qui suit. **Colonne centrale.** Position des résidus sur l'IN (numérotation HXB2), triées par région testée dans les chimères. **Colonnes de droite.** La séquence de l'isolat primaire O utilisé pour les tests est indiquée en jaune dans la première colonne. Alignement de séquences réalisé avec des intégrases d'isolats du VIH-1/O (48 séquences provenant de la base de données LANL et du centre de référence VIH, CHU de Rouen). La séquence consensus est indiquée en noir. Le pourcentage indique le niveau de conservation du consensus au sein du groupe O.

a. Impact des mutants sur le taux de clivage du précurseur Gag

Afin d'évaluer l'implication de chaque résidu dans la perte d'activité, nous avons remplacé un à un ces acides aminés dans l'IN A, par ceux présents chez l'IN O, puis les intégrases mutantes ont été testées dans notre système expérimental pour quantifier l'efficacité de maturation de Gag, l'efficacité de transcription inverse et l'efficacité d'intégration.

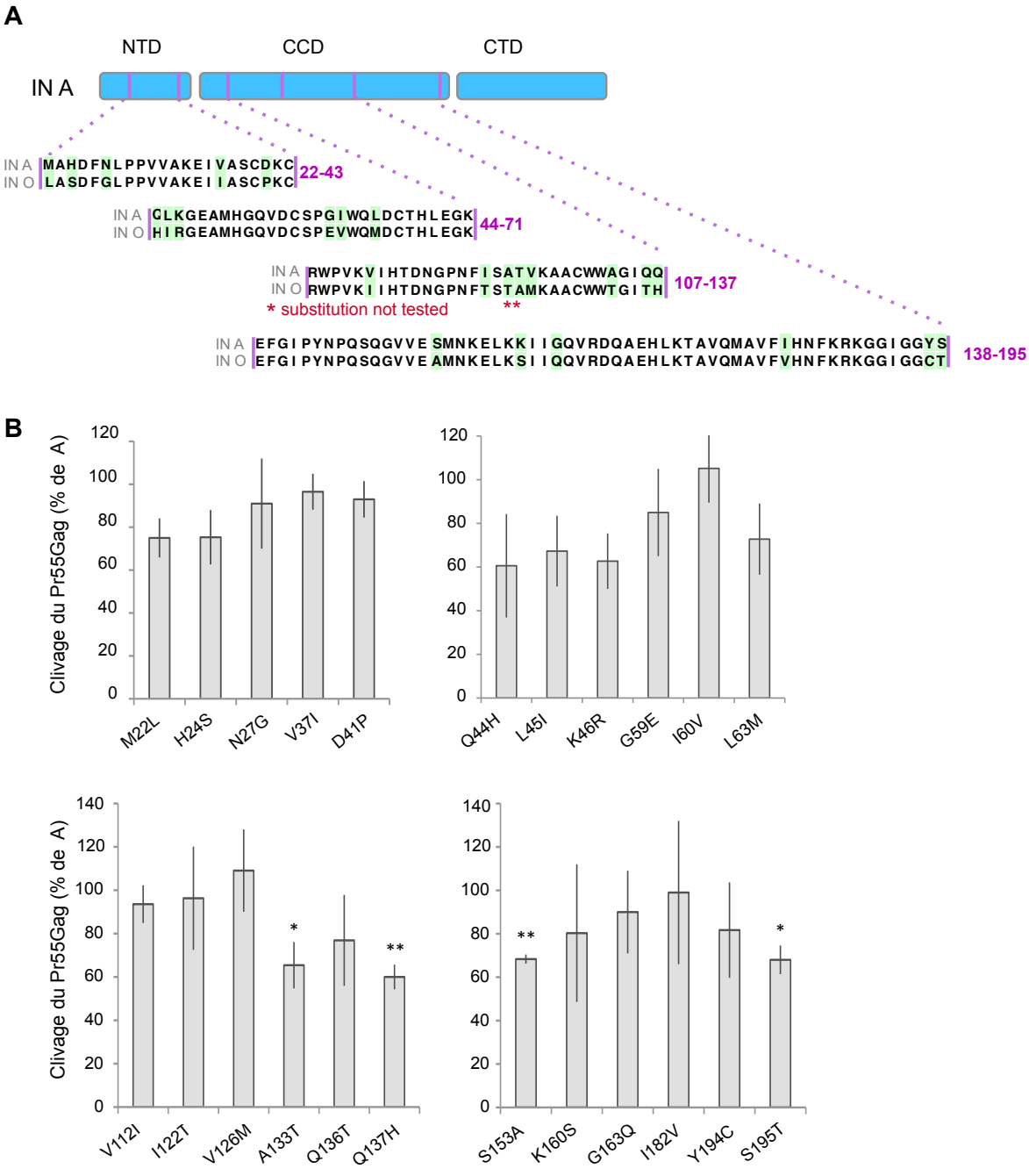


Figure 34 : Taux de clivage protéolytique du précurseur Gag des virus portant les intégrases mutantes. A. Schématisation des régions testées (en violet) sur l'IN parentale A (numérotation HXB2). L'alignement de la séquence de l'isolat A et O indique (en vert) les substitutions qui ont été testées individuellement. **B.** Taux de clivage du précurseur Gag déterminé par quantification des Western blot avec l'anticorps anti-CA. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé à la valeur du parent (A=100%). Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types.
(p<0,05 ; ** p<0,01 ; *** p<0,0001)*

Cinq intégrases mutantes ont été testées pour la région 22-43, six pour la région 44-71 et six encore pour la région 138-195 (Figure **34.A**). Concernant la région 107-137, l'alignement de séquence entre les intégrases de l'isolat A et de l'isolat O montre huit résidus différents. Cependant, l'on remarque que les positions 124 et 125 possèdent les mêmes acides aminés mais avec un ordre inversé (tableau **5**, AT et TA). L'analyse de la conservation des résidus présents à ces positions au sein des groupes M et O montre deux alanines conservées (AA) pour le groupe M et une alanine suivie d'une tyrosine (AT) pour le groupe O. D'ailleurs, bien que les mêmes acides aminés soient concernés, l'on remarque que les deux isolats utilisés (A et O) ne présentent pas les même résidus que ceux conservés chez leur groupe respectif (AT pour l'isolat A et TA pour l'isolat O) (Tableau **5**). Ces deux considérations assemblées suggèrent que les résidus 124 et 125 sont probablement échangeables chez les deux isolats, nous avons donc choisi de ne pas tester les substitutions à ces deux positions et générer six intégrases mutantes pour tester cette région. Chaque intégrase mutante est nommée par la position du résidu, encadrée, à gauche par l'acide aminé de l'IN A et à droite, par celui qui le substitut, présent dans l'IN O (Figure **34.A**).

Les virus portant les différentes intégrases mutantes présentées ci-dessus ont été testées dans notre système expérimental pour l'efficacité de maturation, de transcription inverse et d'intégration. Concernant le taux de clivage du Pr55Gag qui, pour rappel, est le taux de capside mature comparé à la quantité totale de capsides (matures et immatures), la plupart des intégrases testées génèrent une activité quasi parentale. Seul quatre mutants, A133T, Q137H, S153A et S195T présentent un taux de clivage significativement différent du parent A, de l'ordre de 60% (Figure **34.B**).

Ces résultats suggèrent que ces quatre résidus pourraient être impliqués dans le maintien de la dimérisation fonctionnelle du précurseur Gag-Pol M, aboutissant à la maturation de la particule virale.

b. Effets des mutants sur la transcription inverse

Concernant l'efficacité de transcription inverse des défauts marqués sont observés pour certains des mutants générés, en accord avec le faible taux de transcription inverse retrouvé lorsque les régions entières testées sont d'origine O (chimère 22/43, 44/71, 107/137 et 138/195, Figure **33.C**).

La substitution par l'acide aminé présent dans l'IN O des résidus D_{41}, L_{45}, V_{112}, Q_{136}, S_{153}, I_{182}, Y_{194} et S_{195} provoquent une diminution modérée, significativement différente du parent A (entre 50% et 60% de l'activité de transcription inverse de A) alors que, lorsque ce sont les

résidus H_{24}, N_{27} et A_{133} qui sont substitués, le défaut de transcription inverse est plus important (environ 40% de l'activité de A) (Figure **35**).

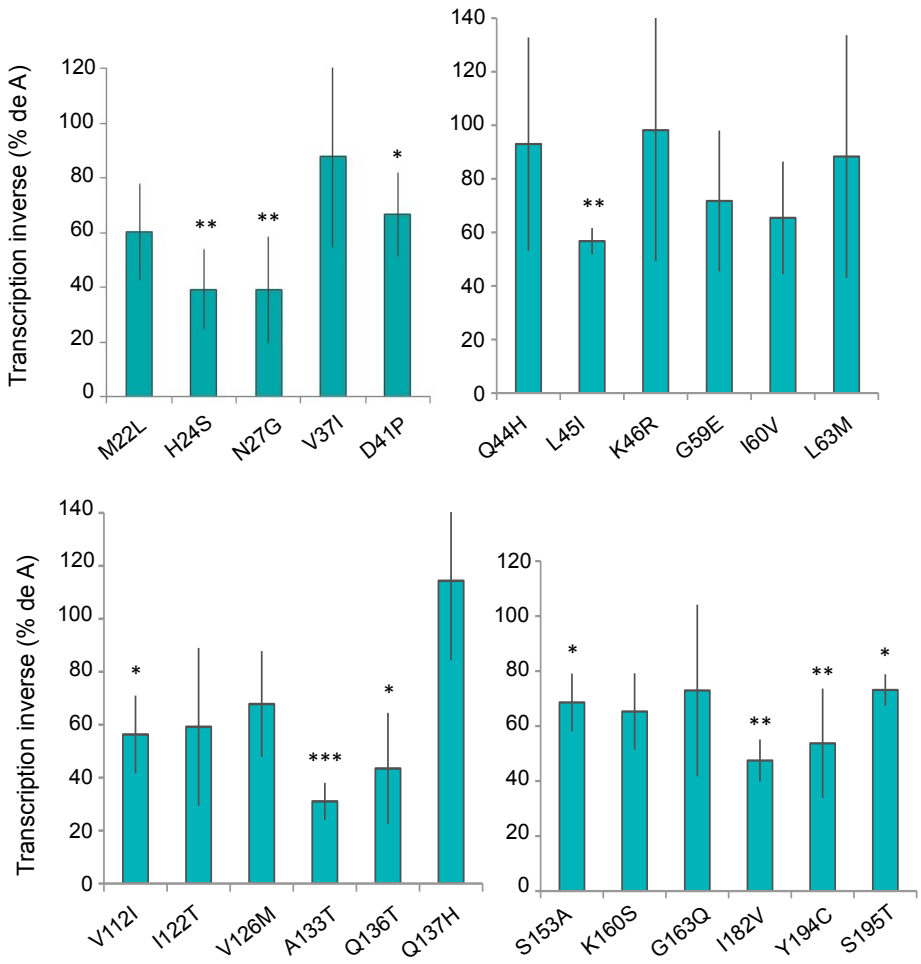


Figure 35 : Transcription inverse des virus portant les intégrases mutantes. Efficacité de transcription inverse (quantité d'ADN viral tardif) comparé au parent A, fixé à 100%. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé à la valeur du parent A. Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types.
(p<0,05 ; ** p<0,01 ; *** p<0,0001)*

Comme les mutants S153A et S195T montrent une diminution de l'activité de transcription inverse similaire à celle du taux de clivage de Gag (Figure **34.B**), on peut supposer que ces baisses d'activités pourraient être liées. A l'inverse, le défaut important du mutant A133T ne peut pas être expliqué par la baisse modérée du taux de clivage.

La comparaison entre les moyennes des quantités de produits précoces et tardifs de la transcription inverse pour chaque mutant ne montre aucune différence significative (Figure **36**), même pour les mutants avec un défaut important de transcription inverse (H24S, N27G et A133T), suggérant que le défaut d'activité est global et pourrait survenir avant la synthèse du premier fragment d'ADN de la réaction, le minus strand strong-stop DNA (-sssDNA) (voir Introduction, partie **I.3.a**).

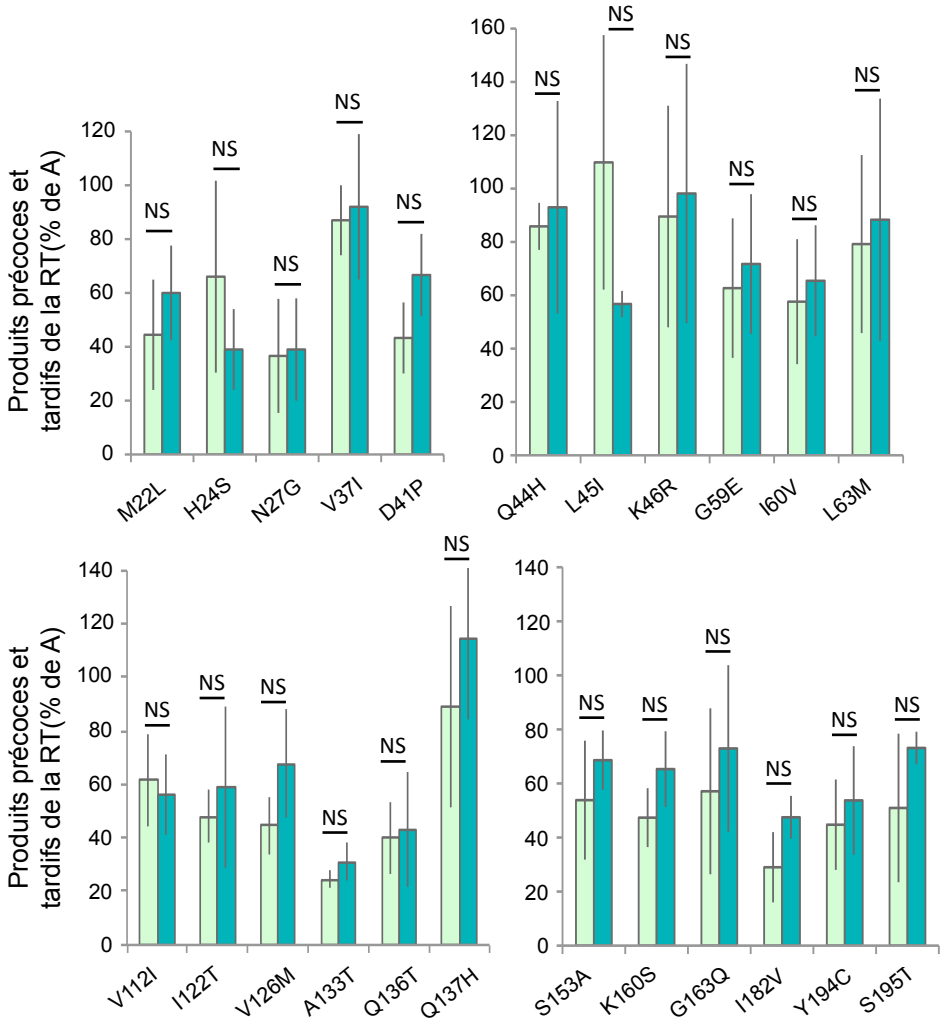


Figure 36: Produits de la transcription inverse. Comparaison entre les quantités des produits précoces (en vert clair) et tardifs (en vert foncé) de la transcription inverse, exprimées en fonction du parent A, fixé à 100%. Un test de Student à deux échantillons indique si les moyennes sont statistiquement différentes pour chaque chimère. Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types.
*(NS p>0,05 ; * p<0,05 ; ** p<0,01 ; *** p<0,0001)*

Ces résultats mettent en évidence des résidus non conservés du NTD et du CCD de l'IN A importants pour la synthèse de l'ADN viral, probablement par le biais du maintien direct ou indirect de l'interaction entre l'IN et la RT.

c. Implications des résidus non conservés dans l'intégration

Comme le taux d'intégration est dépendant de la quantité d'ADN viral disponible, les données d'efficacité d'intégration sont normalisées par la quantité de produits tardifs de transcription inverse. Cette normalisation permet de mieux visualiser si les mutants de l'IN ont un impact sur la transcription inverse et l'intégration ou seulement sur la transcription inverse.

La baisse de la fonctionnalité de la transcriptase inverse n'est pas le seul défaut que présentent certains de ces mutants, puisque de fortes différences sont également observées au niveau de l'intégration. L'intégrase mutante M22L présente un taux d'intégration diminué de moitié comparé au parent A alors que les remplacements des trois autres résidus par l'acide aminé présent dans l'IN O engendrent des intégrases mutantes quasiment inactives (Figure **37**).

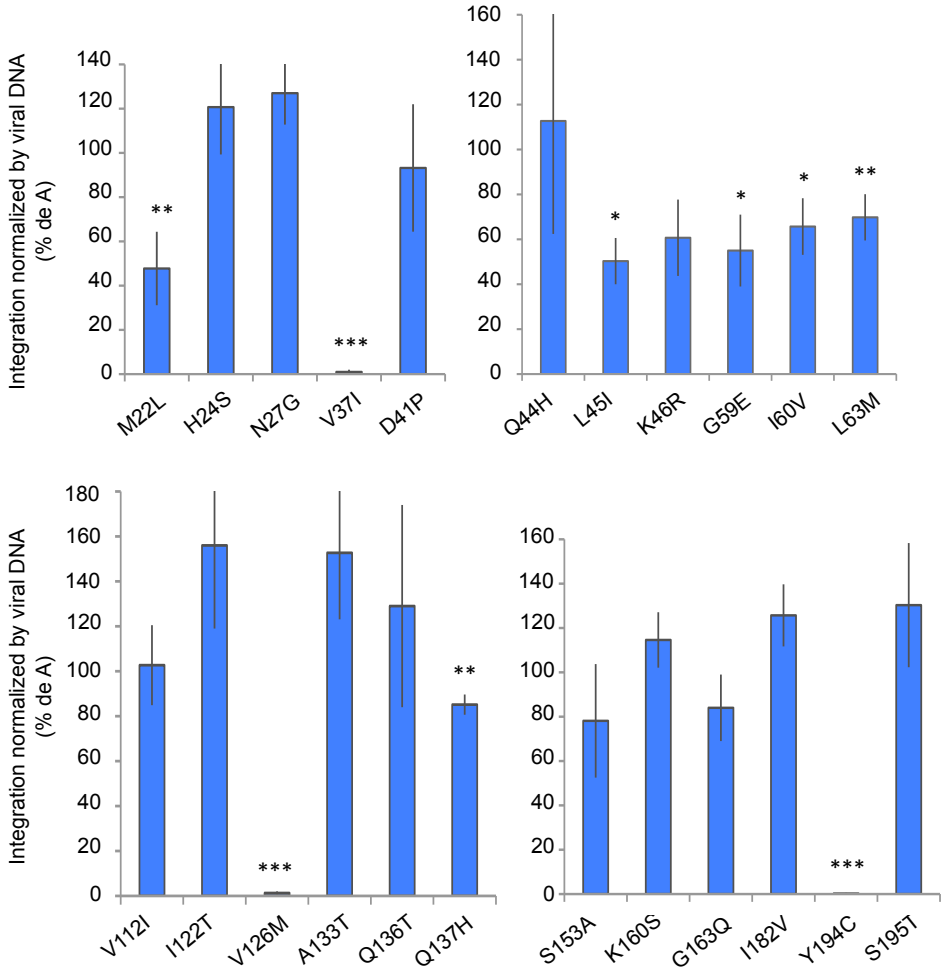


Figure 37 : Efficacité d'intégration des virus portant les intégrases mutantes. Efficacité d'intégration, normalisée par la quantité totale d'ADN viral et exprimée par rapport au parent A, fixé à 100%. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé au parent A. Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types.
(p<0,05 ; ** p<0,01 ; *** p<0,0001)*

Malgré le taux parental observé avec la chimère 44/71, plusieurs mutants dans cette région présentent une diminution modérée mais significative de l'efficacité d'intégration normalisée par la quantité d'ADN viral produit (L45I, G59E, I60V, L63M). Le taux d'intégration est de l'ordre de 50% pour les mutants L45I et G59E, alors que les mutants I60V et L63M présentent une activité résiduelle d'environ 70%. (Figure **37**).

Ces résultats mettent en évidence le rôle majeur des résidu V_{37}, V_{126} et Y_{194} de l'intégrase A dans son activité.

d. Recherche du partenaire coévolutif du résidu 194

Comme décrit dans la partie **I.5.c** de l'introduction, un réseau de coévolution comprend en général plusieurs résidus qui exercent une pression sélective l'un sur l'autre pour conserver leur interaction ou leur compatibilité. Etant donné que l'absence d'intégration de la chimère 138/195 (Figure **33.E**) peut être attribuée à la seule mutation Y194C (Figure **37**), nous avons donc choisi de chercher le partenaire du résidu C_{194} en premier lieu. La cystéine 194 de l'isolat O utilisé n'est pas conservée au sein du groupe O, le consensus étant une tyrosine, comme pour le groupe M. Ce résidu est donc caractéristique de l'isolat O que nous avons utilisé.
Afin de rechercher un éventuel partenaire coévolutif pour le résidu cystéine en position 194, nous avons substitué la tyrosine 194 de l'IN A par la cystéine, sur des chimères possédant une portion de séquence en amont de la position 194 d'origine O (1/21 à 107/137, Figure **38.A**).

L'idée est que si un ou plusieurs acide(s) aminé(s) d'origine O, qui fonctionnerait(aient) en synergie avec la cystéine 194, se trouve(nt) dans l'une de ces régions, la chimère devrait présenter une restauration de fonctionnalité. Nous avons ciblé ces régions car, au vu de l'absence de fonctionnalité de la chimère 137, qui porte la cystéine 194 avec le NTD et le CCD d'origine A et le reste de la protéine (à partir du résidu 137) d'origine O (Figure **30.C**), son(es) partenaire(s) doi(ven)t probablement se trouver en N-ter de l'IN, en amont de la position 137. A cette fin, nous avons utilisé les chimères avec des portions d'IN O plus petites (chimères présentées dans la Figure **33.A**), qui présentent une maturation efficace car la plupart des chimères initiales dans cette région (chimères 1 à 71, Figure **28**) ne sont pas maturées rendant la caractérisation des variations dans l'intégration impossible.

On n'observe aucune restauration de fonctionnalité (Figure **38.B**) comparée à l'intégrase mutante Y194C, les différentes régions remplacées ne présentent donc probablement pas d'acides aminés en « dialogue » avec la cystéine 194. Il est néanmoins possible que le partenaire de la cystéine 194 soit présent dans l'une des régions remplacées mais que d'autres résidus présents dans cette région perturbent la fonctionnalité pour d'autres raisons, en masquant l'effet positif du résidu d'intérêt.

En outre, lorsque l'on remplace la cystéine 194 par la tyrosine dans le contexte de la chimère 138/195, seule une restauration partielle de l'activité (38%) est observée (Figure **38.B**). Ainsi,

bien que la substitution Y194C génère une efficacité d'intégration quasi nulle, le résidu 194 à lui seul n'est pas responsable de la baisse d'activité de la chimère 138/195, suggérant que d'autres résidus non conservés entre les isolats A et O affectent l'efficacité d'intégration, lorsqu'ils sont présents simultanément dans cette région.

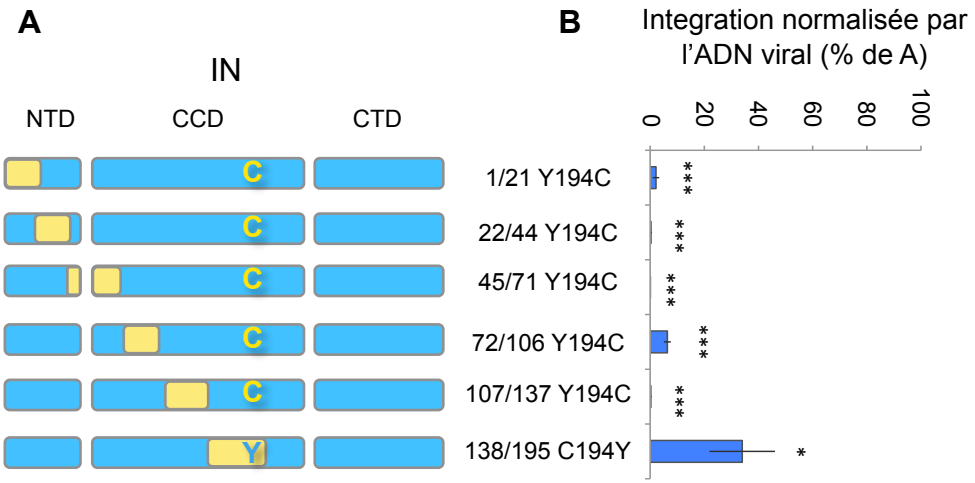


Figure 38 : Test de restauration du mutant Y194C. A. Schéma des IN chimères testées. Les positions des extrémités encadrant la région substituée par la séquence d'origine O sont indiquées à droite du schéma. L'acide aminé présent à la position 194 est indiqué, en bleu pour la tyrosine d'origine A et en jaune pour la cystéine d'origine O. **B.** Efficacité d'intégration normalisée par la quantité totale d'ADN viral et exprimée par rapport au parent A, fixé à 100%. Les étoiles reflètent la p valeur d'un test de Student à un échantillon comparé à A. Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types.
(p<0,05 ; ** p<0,01 ; *** p<0,0001)*

En conclusion, l'analyse des mutants ponctuels de chacune des régions présentant un défaut d'activité a permis de mettre en évidence plusieurs résidus responsables de la chute de fonctionnalité, au niveau de l'intégration ou d'autres étapes du cycle dans lesquelles l'IN joue un rôle. Ces résultats montrent, pour la première fois, que des résidus non conservés entre les groupes sont impliqués dans l'activité de l'intégrase.

2. Caractérisation fonctionnelle des défauts d'intégration

Les résultats précédents ont montré que certaines substitutions de l'intégrase (V37I, V126M et Y194C) génèrent des virus non infectieux, dû à une absence d'intégration. Afin de caractériser le défaut fonctionnel de ces mutants, les différentes formes de cercles à 2LTR (2LTRc) ont été quantifiées. Les quantités de 2LTRc peuvent être informatives d'éventuels défauts dans l'import nucléaire et dans la première étape de l'intégration.

a. Rôle dans l'import nucléaire

Les 2LTRc ont été quantifiés pour chacun des trois mutants de l'intégrase et rapportés à la quantité totale d'ADN viral. Les données sont exprimées en fonction de l'intégrase A catalytiquement inactive (mutation D116A), appelée IN(-) et fixée à 100%, puisque l'ADN viral est importé dans le noyaux mais présent uniquement sous les formes non intégrées (2LTRc, 1LTRc, linéaire). Ainsi, un taux de 2LTRc similaire à l'IN(-) pour les mutants défectueux dans l'intégration sera le signe d'un import nucléaire efficace.

Les deux parents A et O présentent environ 20% de la quantité de 2LTRc de l'IN(-), différence attendue puisqu'ils ont une activité d'intégration sauvage, car la plupart de l'ADN viral est sous forme provirale, intégré au génome de l'hôte. Pour ce qui concerne les trois mutants, on remarque qu'ils possèdent significativement plus de 2LTRc que le parent A, 48%, 50% et 67% pour le mutant V37I, V126M et Y194C respectivement (Figure **39.A**). Cependant le taux de 2LTRc de ces mutants n'atteint pas les 100% attendu pour une intégrase catalytiquement inactive, comme l'IN(-). Ainsi, les données montrent que les mutants de l'intégrase génèrent plus de cercles à 2LTR que le parent A, mais moins que l'N(-).

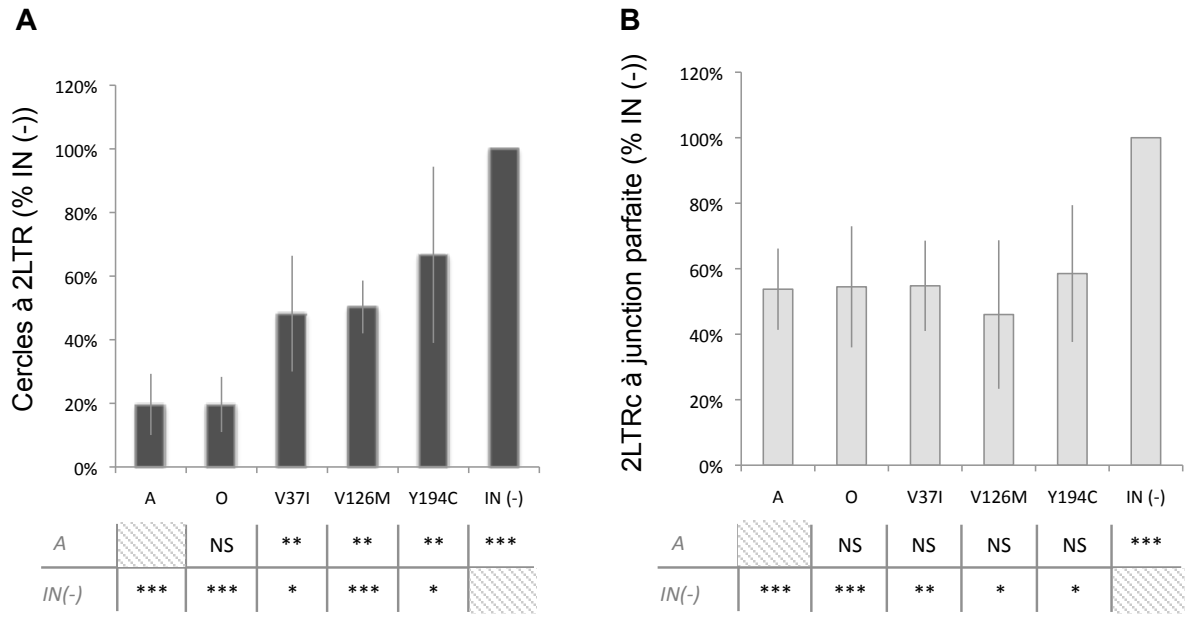


Figure 39 : Caractérisation fonctionnelle des mutants ponctuels. A. Quantités totales de cercles à 2LTRs détectées, exprimées par rapport à l'IN (-), fixé à 100. L'IN (-) est l'IN A dépourvue de son activité catalytique car mutée dans le site actif (mutation D116A). Le tableau en dessous du graphique contient les symboles qui reflètent la p valeur d'un test de Student comparé au parent A (ligne du haut) ou à l'IN(-) (ligne du bas). **B.** Rapports entre les quantités de cercles à 2LTRs à jonctions parfaites et les cercles à 2LTRs totaux, exprimés par rapport à l'IN (-), qui est fixée à 100 %. Le tableau en dessous du graphique contient les symboles qui reflètent la p valeur d'un test de Student comparé au parent A (ligne du haut) ou à l'IN(-) (ligne du bas). Les expériences ont été répétées au moins 3 fois, les barres d'erreurs correspondent aux écart-types.
*(NS p>0,05 ; * p<0,05 ; ** p<0,01 ; *** p<0,0001)*

Ces résultats suggèrent que près de la moitié de l'ADN viral qui n'est pas intégré n'est pas transformé en 2LTRc. Etant donné que l'ADN viral est transformé en 2LTRc par la cellule dans le noyau (système nucléaire de liaison des extrémités non homologues), on peut supposer que la diminution modérée de ces formes (comparé à l'IN (-)) est due à la baisse de la quantité d'ADN viral linéaire dans le noyau, suggérant une baisse d'efficacité d'import nucléaire. Cependant, cette baisse d'activité ne peut expliquer les défauts d'intégration sévères observés pour ces mutants, suggérant la présence d'un autre défaut fonctionnel se situant probablement à une autre étape du processus d'intégration de l'ADN viral dans le génome de l'hôte.

b. Rôle dans l'intégration

La réaction d'intégration se produit en deux étapes, le clivage en 3' des LTRs puis le transfert de brin. Les défauts dans la première étape de la réaction intégration peuvent être détectés par la quantification des 2LTRc à jonctions palindromiques parfaites. En effet les 2LTRc étant formés à partir de l'ADN linéaire, les extrémités de celui-ci peuvent avoir été clivées ou non par l'IN. Ainsi, l'augmentation des 2LTRc à jonctions palindromiques parfaites est signe d'un défaut dans le clivage en 3' des LTRs. Les données sont exprimées par rapport à la quantité totale de 2LTRc, en fonction de l'IN (-), fixée à 100%, puisque cette intégrase ne catalyse pas cette réaction. Si les mutants présentent un taux de 2LTRc à jonctions parfaites similaire à celui de l'IN (-) et, donc, différent du parent A, ce sera le signe d'un défaut dans le clivage en 3' des LTRs.

Les deux parents A et O sont à environ 50% de l'activité de l'IN(-), différence attendue puisqu'ils ont une activité catalytique sauvage, la moitié des 2LTRc ont une jonction palindromique imparfaite car l'ADN viral a été clivé par l'IN (Figure **39.B**). Chacun des trois mutants testés présente une activité similaire au parent A et significativement plus basse comparée au taux de 2LTRc à jonction parfaites de l'IN(-). Ces résultats suggèrent que les substitutions V37I, V126M et Y194C dans l'intégrase, qui affectent l'intégration, ne perturbent pas l'activité de clivage en 3' des LTRs.

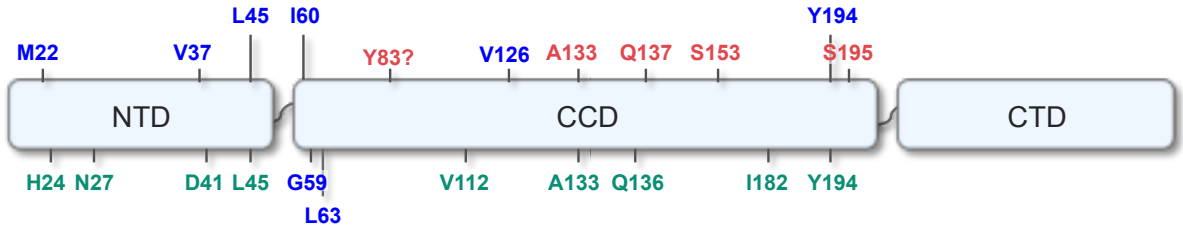
En conclusion, ces résultats ont montré que les mutants V37I, V126M et Y194C ont un impact modéré sur l'import nucléaire et présentent probablement un défaut à une autre étape nécessaire à l'intégration (transfert de brin, transitions conformationnelles du tétramère de l'IN pendant le processus d'intégration).

Discussion

1. Impact de l'IN sur la maturation

Grâce à la construction de chimères le long de l'intégrase entre les isolats A et O, nous avons pu mettre en évidence la présence de réseaux de coévolution jouant différents rôles dans le cycle infectieux. Malgré le fait que nous ayons choisi, dans notre étude, de caractériser prioritairement les chimères présentant des défauts de transcription inverse et d'intégration, nous avons aussi observé une baisse du taux de clivage du précurseur Pr55Gag avec certaines de ces chimères suggérant la présence de résidus non conservés entre A et O, permettant le maintien de la maturation des particules virales. Parmi les régions étudiées, plusieurs présentent une baisse du taux de clivage de Gag (44/71, 107/137 et 138/195). L'analyse individuelle des acides aminés qui diffèrent entre les isolats A et O a mis en évidence quatre résidus (Figure **40**) jouant un rôle dans le clivage des précurseurs, corrélée avec l'activité observée lorsque les régions entières sont substituées par la séquence provenant du O.

A



B

Fonctionnalité (60% - 40%)	Fonctionnalité (40% - 20%)	Fonctionnalité (20% - 0%)
A133 ; Q137 ; S153 ; S195		
D41 ; L45 ; V112 ; Q136 ; Y194	H24 ; N27 ; A133	
I60 ; L63 ; M22 ; L45 ; G59		V37; V126 ; Y194

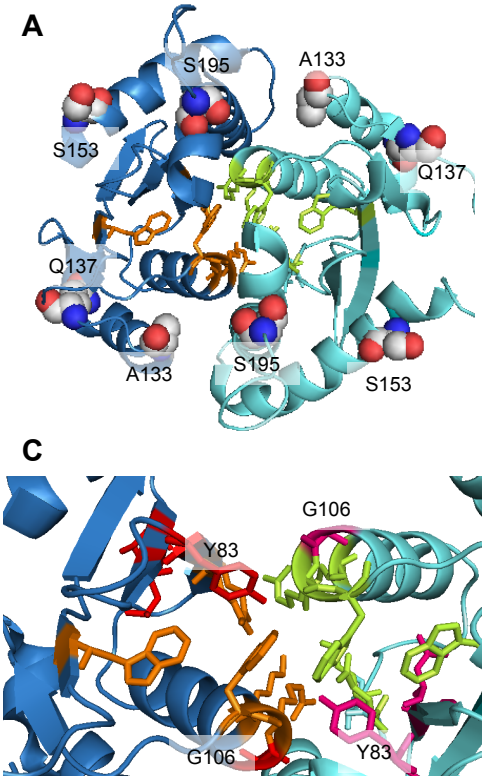
Clivage du Pr55Gag
Transcription inverse
Integration

Figure 40 : Récapitulatif des résultats. A. Schéma des résidus non conservés de l'IN impliqués dans le maintien de sa fonctionnalité (maturation, en rouge ; transcription inverse, en vert, integration, en bleu). **B.** Tableau indiquant ces mêmes résidus classés par l'activité observée lorsqu'ils sont substitués par l'acide aminé de l'IN O.

Concernant la région 44-71, aucune des substitutions individuelles ne présente un défaut de maturation bien que lorsqu'elles sont présentes simultanément dans la chimère, le clivage du précurseur Gag est affecté. Comme cette région contient la partie flexible qui relie le NTD et

le CCD ainsi que le début du CCD, on peut supposer que le remplacement par la séquence d'origine O peut perturber son positionnement et donc la structure du précurseur Gag-Pol. Ces résultats sont en concordance avec l'observation précédente soulevant l'idée que la présence simultanée des résidus O dans le CCD de l'IN, au sein du précurseur Gag-Pol M, perturbe la maturation.

Le résidu Q_{137} a précédemment été identifié dans la littérature[319] par l'étude de la mutation Q137R, qui présente un sévère défaut d'infectivité dû à l'impact sur de nombreuses fonctionnalités de l'IN (défaut de liaison avec INI1, défaut de production des produits précoces et tardifs de la transcription inverse, mauvaise interaction avec LEDGF et RT, morphologie aberrante de la particule).



B

Sous-type A groupe M	Isolat A	Position du résidu	Isolat O	Groupe O
		72-106		
V 87%	V	72	I	I 55%
L 97%	L	74	I	I 100%
Y 100%	Y	83	F	F 100%
A 99%	T	91	A	A 98%
E 99%	D	96	E	E 100%
L 87%	I	101	L	L 100%
G 95%	G	106	A	A 96%

Figure 41 : A. Structure du dimère de CCD (PDB code : 1EX4), chaque protomère est représenté en nuance de bleu. Les résidus de la surface de dimérisation sont respectivement représentés en orange et en vert. Les quatre résidus (Q_{137},A_{133},S_{153},S_{195}) non conservés sont indiqués. **B.** Conservation de séquences des IN M et O dans la région 72-106. La séquence consensus de chaque groupe est indiquée en noir suivi du pourcentage de conservation au sein du groupe. **C.** Agrandissement de la structure de A. Les résidus différents entre les isolats A et O de la région 72-106 sont indiqués en rouge et en rose, par chaque CCD.

Par ce travail, nous avons montré que la présence d'une histidine, contrairement à celle de l'arginine, perturbait peu l'infectivité (activité de transcription inverse et d'intégration parentale), suggérant que ce n'est pas seulement la présence d'une charge positive en position 137 qui serait responsable de la perte d'activité mais plutôt la taille et la structure de la chaine latérale. Ainsi, comme le CCD est le siège de nombreuses interactions (surface de dimérisation, interaction avec LEDGF), on peut se demander si le résidu Q_{137} participe à des interactions fonctionnelles de l'IN qui ne sont pas maintenues en présence du groupe guanidinium de l'arginine mais le sont en présence du groupe imidazole de l'histidine.

D'ailleurs, en accord avec nos résultats, la substitution Q137H de l'IN a été démontrée générer des virus infectieux en cycle unique et multiple[320]. Cependant, l'histidine n'est pas retrouvée dans la pandémie, puisque le résidu Q_{137} est conservé à 99% au sein du groupe M[321] suggérant qu'il aurait un rôle *in vivo* non révélé par les tests d'infectivité sur cellules en culture.

Par la suite, il serait intéressant d'identifier le rôle précis de chaque résidus/régions identifié et de caractériser quels acides aminés de la région 72-106 sont importants pour le clivage des précurseurs. D'autant plus que la mise en évidence des résidus identifiés (A_{133}, Q_{137}, S_{153} et S_{195}) dans la structure du dimère de CCD-CTD (PDB code : 1EX4) montre qu'ils encadrent la surface de dimérisation du CCD (conservée chez les Lentivirus, résidus $W_{61}E_{85}E_{87}K_{103}R_{107}W_{108}$)[223], qui pourrait aussi jouer un rôle dans la dimérisation des précurseurs Gag-Pol (Figure **41.A**). Il se pourrait que ces quatre acides aminés perturbent la structure de la surface de dimérisation lorsqu'ils sont mutés, affectant ainsi la maturation.
D'ailleurs, sur cette même structure nous avons mis en évidence les résidus de la région 72-106 différents entre les isolats A et O, et conservés au sein de leur groupe respectif. Parmi les sept résidus différents, quatre résidus présentent ces critères (V_{72}, L_{74}, Y_{83} et G_{106}), car les trois autres (A_{91}, E_{96}, L_{101}) sont conservés entre le sous-type A du groupe M et le groupe O mais typiques de l'isolat A (Figure **41.B**). On remarque que Y_{83} et G_{106} sont à l'intérieur de la surface de dimérisation (Figure **41.C**). De plus, le groupement hydroxyle de la tyrosine 83 de chaque protomère pointe vers le centre de cette surface, suggérant qu'ils pourraient établir une(des) interaction(s) dans la surface de dimérisation. La substitution de la tyrosine 83 en phénylalanine (résidu présent dans l'IN O) briserait cette(ces) interaction(s) puisque la phénylalanine, bien qu'elle présente une structure similaire à la tyrosine, n'a pas de groupement hydroxyle dans sa chaine latérale.

Ainsi, nous avons montré l'implication de l'IN dans la maturation, par le biais d'acides aminés non conservés impliqués dans des interactions nécessaires au maintien de la dimérisation fonctionnelles des précurseurs Gag-Pol, permettant l'autoclivage et, donc, l'activation de la protéase mature.

2. Effets de l'IN sur la transcription inverse

Nous avons également pu constater une diminution de la quantité d'ADN viral produit par transcription inverse pour certains virus portant les IN chimères. Ainsi, malgré la conservation de la surface d'interaction fonctionnelle avec la RT au sein du CTD, des résidus

non conservés du NTD et/ou du CCD doivent être impliqués directement ou indirectement dans le maintien de cette interaction permettant le bon déroulement de la transcription inverse. L'analyse des résidus qui diffèrent entre les isolats A et O, pour chacune des régions présentant une baisse importante de la quantité de produit final de la transcription inverse (22-43, 44-71, 107-137 et 138-195) a permis de mettre en évidence plusieurs acides aminés de l'IN plus ou moins essentiels pour une transcription inverse efficace (Figure **40**).

Il est intéressant de souligner que les résultats des substitutions analysées individuellement dans la région 44-71 ne sont pas en accord avec l'analyse de la transcription inverse lorsque toute la région est d'origine O, démontrant un effet cumulatif des défauts opérés par les substitutions. Ainsi, les résidus non conservés de cette région semblent jouer, tous ensemble, un rôle dans la RT non retrouvé lorsqu'ils sont individuellement substitués par l'acide aminé provenant de l'IN O.

Nos résultats indiquent que le défaut de RT pour toutes ces substitutions concernerait les phases précoces de la synthèse d'ADN, suggérant qu'ils affectent l'interaction RT-IN, nécessaire pour l'initiation de la transcription inverse[272-275]. Il se pourrait que ces mutations affectent la conformation de la protéine, altérant l'interaction IN-RT, puisque les interactions inter domaines entre le CCD et le CTD, et entre le CTD et la boucle flexible reliant le NTD au CCD décrites à partir de la structure du tétramère[58] (Figure **17.C**) laissent supposer qu'elles sont nécessaires au maintien de la structure de la forme fonctionnelle de l'IN. En effet, la perturbation de la structure de l'enzyme peut changer l'exposition de la surface d'interaction du CTD, la rendant moins accessible pour la transcriptase inverse. D'ailleurs, la mutation d'un résidu, C130S de l'IN, à proximité de la position 133, affecte l'interaction IN-RT puisqu'elle résulte en l'abolition de l'activité de la transcriptase inverse[322]. On peut supposer que la présence d'un résidu polaire avec un groupement hydroxyle, la thréonine (en position 133) à la place de l'alanine non polaire, proche de la C_{130} (résidu polaire avec un groupement thiol), peut avoir un effet similaire à celui de son remplacement par la sérine (C130S), un autre acide aminé polaire avec un groupement hydroxyle.

La diminution de l'efficacité de transcription inverse observée avec la substitution des résidus S_{153} et S_{195} par les acides aminés provenant de l'IN O est corrélée avec la diminution du taux de clivage des précurseurs, suggérant que celles-ci sont directement reliées : il y a moins d'activité de transcription inverse puisqu'il y a moins de RT mature. On ne peut cependant pas exclure un éventuel défaut d'activité de la transcriptase inverse dans le cas de ces deux substitutions de l'IN. En effet, la RT étant en excès dans la particule virale (50-100 molécules par virions pour 2 ARN génomiques), le défaut modéré dans la maturation pourrait être compensé et la baisse de transcription observée serait donc, dans ce cas, due à un défaut de transcription inverse. Le résidu A_{133} présente également un défaut de maturation lorsqu'il est substitué par l'acide aminé de l'IN O, cependant la chute sévère de l'efficacité de

transcription inverse lorsqu'il est muté laisse à présager un défaut plus important dans la synthèse de l'ADN viral.

Cette étude révèle que des acides aminés non conservés entre différents groupes et types de VIH sont impliqués dans la stabilisation de l'interaction IN-RT nécessaire pour une transcription inverse efficace. Ces acides aminés ne sont pas localisés dans le CTD, où se situe la surface d'interaction avec la RT ($R_{231}W_{243}G_{247}A_{248}V_{250}I_{251}K_{258}$)[273,274] conservée chez le VIH, mais dans d'autres régions de la protéine qui affectent indirectement le maintien de l'interaction entre la RT et l'IN, probablement par la déstabilisation de la structure de l'IN compatible avec le maintien de la transcription inverse. Ainsi, nous avons mis en évidence la coévolution inter protéique entre la RT et l'IN, puisque des acides aminés non conservés de l'IN sont nécessaires au bon fonctionnement de la RT.

3. Impact des mutants sur l'intégration

Enfin, en plus des défauts de maturation et de transcription inverse, nous avons également observé de sévères diminutions de l'efficacité d'intégration avec certaines chimères, suggérant la présence de réseaux de coévolution permettant de maintenir l'activité catalytique de l'IN.
Malgré le taux parental d'intégration lorsque la région 44-71 est d'origine O dans l'IN A (analysée finement puisqu'elle présente une baisse importante de la RT), plusieurs substitutions individuelles des résidus de cette région présentent une diminution de près de la moitié de l'efficacité d'intégration parentale (résidus L_{45}, G_{59}, I_{60} et L_{63}) (Figure **40**). Ces résidus font partie de la boucle flexible reliant le NTD au CCD, dont plusieurs positions à proximité (K_{42}, E_{48}, Q_{53}, C_{56}) sont supposées être en interaction inter domaines avec le CCD et le CTD (Figure **17.C**)[58]. Il est possible que les mutations ponctuelles perturbent la structure de cette boucle, affectant ainsi ces interactions nécessaires pour la structuration de l'intasome qui permet une intégration optimale. Or lorsque l'ensemble de ces résidus est d'origine O, les mutations peuvent se compenser et permettre d'adopter une conformation favorable à ces interactions.

L'analyse des résidus qui diffèrent entre les isolats A et O pour les régions 22-43, 107-137 et 138-195 a permis de mettre en évidence plusieurs acide aminés isolés de l'IN importants pour l'intégration en elle-même (Figure **40**). Les résidus M_{22}, V_{37}, V_{126} et Y_{194} affectent significativement l'efficacité d'intégration lorsqu'ils sont substitués par le résidu d'origine O dans une IN A. Parmi ces résidus, certains affectent drastiquement l'intégration lorsqu'ils sont mutés (V_{37}, V_{126} et Y_{194}). Nous avons procédé à une caractérisation fonctionnelle de ces

différentes substitutions en analysant les formes non intégrées de l'ADN viral, informatives de l'efficacité d'import nucléaire et du clivage en 3' des LTRs. En effet, une chute de l'intégration (en dehors d'un défaut dans la synthèse de l'ADN viral) peut être due à un mauvais import nucléaire ou à un défaut d'une des étapes de la réaction d'intégration (clivage des LTRs ou transfert de brin). Les résultats ont montré que les mutants V37I, V126M et Y194C ont un impact modéré sur l'import nucléaire et une activité de clivage des LTRs parentale. Les résidus V_{37}, V_{126} et Y_{194} ont donc probablement un rôle majeur dans une autre étape nécessaire à l'intégration (transfert de brin, transitions conformationnelles du tétramère de l'IN pendant le processus d'intégration).

Le résidu M_{22} se trouve à côté de résidus impliqués dans des interactions inter domaines (N_{18} et R_{20} en interaction avec K_{188}, Figure **17.C**)[58] identifiés dans le complexe tétramérique de transfert de brin (voir introduction, partie **II.2.c**). Cette lysine 188 fait partie de la boucle entre les hélices α5 et α6 du CCD, boucle qui adopte une conformation contrainte par de nombreux ponts hydrogènes avec un NTD intermoléculaire (en particulier un pont salin avec le résidu N_{25}) et participe à l'interface permettant la formation du tétramère formée par le NTD et le CCD. De façon étonnante, nous avons également identifié un autre résidu de la boucle entre les hélices α5 et α6 du CCD pour son implication dans l'intégration, le résidu Y_{194}. Comme cette boucle a été décrite pour contribuer à l'interface du tétramère d'IN[210], les substitutions M22L et Y194C pourraient perturber les interactions nécessaires à sa multimérisation.

D'ailleurs le résidu Y_{194} a précédemment été identifié pour son rôle essentiel dans la multimérisation de l'IN, aboutissant à la formation du tétramère. Une analyse par chromatographie à exclusion de taille a mis en évidence que deux intégrases mutantes à la position 194, se comportaient de façon diamétralement opposée par rapport à la formation du tétramère : le mutant Y194E qui est majoritairement présent sous forme de monomère, et le mutant Y194F, qui comme l'IN sauvage, est présent en majorité sous forme de tétramère[323]. Comme la substitution Y194F ne semble pas perturber l'activité alors que la mutation de la tyrosine 194 en acide glutamique ou en cystéine aboutit à un défaut d'intégration, on peut supposer que la structure de la phénylalanine (très similaire à celle de la tyrosine) n'affecte pas l'intégration, alors que l'apport d'une charge négative avec l'acide glutamique ou d'une structure bien plus petite, avec la cystéine, comme dans notre étude, résulte en une baisse de l'activité de l'IN, probablement dû à un défaut dans la formation du tétramère. Nous supposons donc que la tyrosine n'est pas impliquée dans des interactions spécifiques à la formation du tétramère (autrement la substitution par F, un acide aminé non polaire, aurait rendu l'IN défectueuse) mais qu'elle se situe à une position charnière, où son encombrement stérique est nécessaire pour maintenir les interactions qui permettent la

formation stable du tétramère. Cette hypothèse est en accord avec nos résultats puisque nous avons montré que l'IN Y194C avait une activité de clivage des LTR parentale et, comme le dimère d'IN est suffisant pour cette activité[203], la perturbation du tétramère résulterait plutôt en l'absence de transfert de brin, étape successive au clivage en 3'. La diminution de l'import nucléaire pourrait s'expliquer par la perturbation de la liaison des cofacteurs cellulaires de l'IN (LEDGF, Transportine 2, importines 3 et 7)[276-281,289] (voir introduction, partie **II.3.a**) favorisant cette étape : la structure de l'IN étant différente, les surfaces d'interactions peuvent être exposées différemment, rendant l'IN plus difficilement accessible aux cofacteurs.

Nous avons montré que la substitution Y194C n'était pas la seule responsable de la baisse d'intégration de la chimère 138/195, puisque seule une restauration partielle était observée lorsque la tyrosine 194 est introduite dans la chimère (Figure **38.B**). Comme cette IN chimère présente la boucle, impliquée dans la formation stable du tétramère, d'origine O, on peut supposer que l'ensemble des résidus O n'est pas complètement compatible avec le NTD d'origine A et perturbe le positionnement et les interactions inter-boucles, inhibant leur effet positif sur la multimérisation de l'IN et induisant la baisse d'intégration. D'ailleurs, une analyse *in silico* a montrée, par un modèle d'homologie avec la structure de l'IN de PFV, que la structure générale de l'IN O semblait très similaire à celle de l'IN M, à l'exception de quelques résidus clés, à proximité de la triade catalytique (dont A_{153} fait partie)[324]. On peut donc se demander si la présence simultanée des résidus O dans la portion 138-195 de l'IN A perturbe la structure de la protéine, notamment celle du site catalytique, affectant ainsi son activité.

La définition par Cryo-EM d'une nouvelle structure de l'intégrase correspondant à la fin de l'étape de transfert de brin, représentant un tétramère de tétramères (hexadécamère), a mis en évidence des résidus formant des interactions spécifiques de ce complexe plus imposant que les précédents, dont font partie E_{35} et K_{240}, plus amplement décrits dans la discussion de la partie **III** (motif NKNK). Le résidu V_{37} est à proximité direct de la position 35, on peut donc supposer que sa mutation en isoleucine (structure de la chaîne latérale plus longue d'un carbone) pourrait modifier le positionnement de E_{35} et ainsi perturber son interaction avec le résidu K_{240}, et donc l'activité de l'IN. D'ailleurs, en accord avec nos résultats, la substitution V_{37}E de l'IN génère des virus non infectieux, confirmant l'importance du résidu V_{37} pour la réplication virale[320]. Nos résultats ont montré que la mutation V_{37}I avait un impact modéré sur l'import nucléaire et un impact sévère sur l'intégration mais pas sur le clivage de l'ADN viral. Dans la mesure où le dimère d'IN est suffisant pour assurer dans le cytoplasme un niveau parental de clivage en 3' des LTRs, la mutation V_{37}I ne devrait pas affecter la dimérisation. Cependant, comme l'on suppose qu'elle pourrait affecter la structure du complexe

tétramérique de l'IN, la diminution de l'import nucléaire tout comme l'absence d'intégration concordent avec cette hypothèse. Comme vu en amont, la perturbation de la structure de l'IN pourrait affecter la liaison des cofacteurs importants pour l'import nucléaire, provoquant une baisse d'efficacité de cette étape, alors qu'une mauvaise conformation ou transition conformationnelle du tétramère de l'IN lors du transfert de brin résulterait en l'absence d'intégration.

Le résidu V_{126} se trouve juste à côté du site de liaison au cofacteur de l'intégrase LEDGF/p75 ($A_{128}A_{129}W_{131}W_{132}I_{161}V_{165}R_{166}E_{170}L_{172}K_{173}$)[219,220], il est, donc, possible que la substitution V126M perturbe la conformation de l'hélice dans laquelle ce résidu se situe et modifie la conformation du site de liaison à LEDGF. La liaison de l'intégrase à son cofacteur étant nécessaire pour assurer l'import nucléaire et l'intégration en elle même, les baisses d'activité observées pour l'IN mutante V126M sont corrélées avec cette hypothèse.

En conclusion, l'analyse des mutants ponctuels de chacune des régions présentant un défaut d'activité a permis de mettre en évidence, pour la première fois, plusieurs résidus non conservés responsables de la chute de fonctionnalité au niveau de l'intégration ou d'autres étapes du cycle dans lesquelles l'IN joue un rôle, probablement due, dans la plupart des cas, à la perturbation d'interactions fonctionnelles ou structurelles.

III. Le motif NKNK du CTD

Les résultats précédents ont mis en évidence plusieurs régions de l'IN susceptibles d'être soumises à des contraintes coévolutives, dont un éventuel réseau de coévolution dans le CTD. Ce domaine est très riche en résidus basiques et possède un domaine responsable de la liaison à l'ADN de manière aspécifique ainsi que la surface d'interaction avec la transcriptase inverse. Au vu de ses nombreux rôles pour assurer la catalyse, il serait intéressant d'identifier les contraintes coévolutives du domaine CTD de l'IN, ainsi que leur fonctions respectives. Afin d'identifier les résidus responsables de la perte d'activité et de caractériser leur rôle, nous avons procédé à un analyse fonctionnelle de cette région de l'IN. Cette partie fait l'objet d'un article en préparation : ***Flexible but conserved: a new essential motif in the C-ter domain of integrase characteristic of group M*.**

Résumé de l'article

Afin de déterminer quelles sont les contraintes coévolutives auxquelles est soumis le domaine C-terminal de l'intégrase, nous avons caractérisé la fonctionnalité observée avec des chimères entre deux isolats primaires provenant de groupes phylogénétiques différents (M et O). Le caractère chimérique du CTD génère une baisse d'efficacité de transcription inverse et d'intégration, signe de la présence de résidus non conservés entre les isolats, importants pour la fonctionnalité.

L'analyse plus fine de cette région a permis de mettre en évidence un motif de quatre résidus essentiel à l'intégration, composé de deux lysines alternées avec deux résidus polaires, conservé chez le groupe M ($N_{222}K_{240}N_{254}K_{273}$). En effet le remplacement des deux lysines ou des deux acides aminés polaires dans l'intégrase de l'isolat du groupe M a un impact sévère sur l'efficacité d'intégration. De façon étonnante, le changement de position des deux lysines au sein du motif de quatre résidus ne perturbe pas significativement l'intégration dans la plupart des cas, suggérant que la position de ces lysines est relativement flexible dans le motif. Ainsi, nous avons pu montrer que la présence d'au moins deux lysines est nécessaire pour observer l'intégration, de façon relativement indépendante de leur positions dans le motif. Cependant, comme nous avons observé des divergences dans l'efficacité d'intégration en fonction des positions que les deux lysines occupent, les positions et la nature des acides aminés du motif semblent tout de même être liées. En effet, les quatre positions du motif ne sont pas équivalentes, deux positions (254 et 273) génèrent

une activité nettement plus basse lorsque les lysines y sont présentes comparée aux deux autres positions (222 et 240).

Comme le motif étudié est strictement conservé parmi les isolats du groupe M, nous avons également évalué son importance dans d'autres sous-types du groupe M que celui étudié principalement dans notre travail, le sous-type A. Nous avons ainsi analysé des isolats du sous-type B, C et de la CRF 02, et constaté que l'absence de lysine dans le motif de ces intégrases conduit également à une perte presque totale de la capacité d'intégration. De plus, nous avons conduit une analyse des variants du motif qui sont fonctionnels dans notre test d'infection en cycle unique mais non retrouvés dans la pandémie, avec des virus réplicatifs (collaboration avec C. Moog). Cette analyse a confirmé nos résultats indiquant que ceux-ci ne reflètent pas un biais dû au système d'infection en cycle unique. Ces résultats ont donc permis de définir un motif flexible, essentiel pour la fonctionnalité des intégrases du groupe M. Cependant, malgré le fait qu'une certaine flexibilité de la séquence soit tolérée pour l'intégration, ce motif est strictement conservé parmi les isolats du groupe M dans la pandémie, ce qui suggère que son intégrité et l'arrangement de ses résidus est probablement nécessaire pour des fonctions spécifiques *in vivo*, qui ne sont pas reproduites dans un système d'infection sur cellules en culture.

La compréhension de toutes les fonctions dans lesquelles ce motif est impliqué pendant l'infection apparaît essentielle pour améliorer notre connaissance de la fonctionnalité de l'intégrase. Afin d'identifier quelles étapes du processus d'intégration sont affectées par les mutations du motif NKNK, la caractérisation des formes non intégrées de l'ADN a été entreprise. En effet, celles-ci peuvent être informatives d'éventuels défauts dans l'import nucléaire ou dans la première étape de l'intégration, le clivage en 3' de l'ADN viral. La caractérisation fonctionnelle de ce motif nous a conduit à penser qu'il jouerait un rôle modéré dans l'import nucléaire et un autre, plus important, dans l'intégration et notamment dans la première étape consistant en un clivage en 3' des extrémités de l'ADN viral.

En conclusion, ce travail a permis de mettre en évidence un motif de quatre résidus, essentiel à l'intégration, qui malgré une certaine flexibilité de séquence est strictement conservé dans la pandémie.

Flexible but conserved: a new essential motif in the C-ter domain of integrase characteristic of group M

Marine Kanja[a], Pierre Cappy[a], Nicolas Levy[c], Oyndamola Oladosu[c], Sylvie Schmidt[b], Romain Gasser[a], Paola Rossolillo[a], Claudia Elefante[a], Christiane Moog[b], Marc Ruff[c], Matteo Negroni[a]*, and Daniela Lener [a]*.

[a] *Retroviruses and Molecular Evolution, Architecture et Réactivité de l'ARN, UPR 9002, IBMC, Strasbourg University - CNRS, R. Descartes 15, F-67000 Strasbourg, France;*

[b] *Molecular Immuno-Rhumatology Laboratory, UMR1109, FMTS, Université de Strasbourg, INSERM, Institut de Virologie, 3 rue Koeberlé, Strasbourg, France*

[c] *Chromatin Stability and DNA mobility, Department of Structural Biology and Genomic, IGBMC, Strasbourg University, CNRS, INSERM, L. Fries St.1, 67404 Illkirch, France.*

* Corresponding author: m.negroni@ibmc-cnrs.unistra.fr ; d.lener@ibmc-cnrs.unistra.fr

Mailing address:

Matteo Negroni ; Daniela Lener
Institut de Biologie Moléculaire et Cellulaire
15 rue René Descartes
67084 Strasbourg Cedex, France

Abstract

Structural and functional studies on HIV-1 integrase have been mostly carried out using mutants of laboratory-adapted strains or isolates from subtype B of group M. Here we use a different approach by generating chimerical integrases between HIV-1 primary isolates from groups M and O (groups that have been shown to generate several recombinant forms in nature) and analysing their functionality by infection in culture. By this mean we replace the amino acids of one integrase (group M) with those present at the same position in the other integrase chosen (group O). The rationale is that, if coevolution networks specific to each group are perturbed in the chimeras, functionality might be impaired. The characterisation of the functional defect and the mapping of the regions participating to the coevolution network will be informative of the involvement of these regions in the function altered.

We indeed observe a decrease of integration efficiency for certain M/O chimeras. Systematic replacement of residues that differ between the CTD of wt and chimerical IN defines a motif of four residues, two lysines and two polar amino acids (NKNK for group M), essential for integration. Indeed, in group M, replacement of one or both lysines or of the two polar amino acids has a dramatic impact on integration efficiency. Remarkably we observe that the two K residues do not need to be present in fixed positions of the motif for integration to occur. Indeed, the presence of at least two K is needed to observe integration regardless their position in the motif. However, depending on the positions they occupy, integration does not occur with equivalent efficiency. These results suggest that the positions and the nature of the amino acids of the motif are linked. This, even though some of the residues involved had been reported individually to be essential for integration in previous work.

These results are relevant from at least two standpoints. One is that they define a new, flexible, motif essential for IN functionality. The other is that, despite the flexibility of sequence tolerated for carrying out integration, this motif is strictly conserved among isolates in the pandemic suggesting that its integrity and the ordered arrangement of its residues is required for additional specific functions in vivo. Understanding all the functions in which this motif is involved during infection appears now essential for improving our knowledge of IN functionality.

Acknowledgment: This work was supported by Sidaction (grant No. AI25-1-02335). Marine Kanja is the recipient of a Sidaction Ph. D. fellowship No.BI25-1-02305.

Introduction

HIV integrase (IN) is one of three enzymes that guarantee viral replication as it integrates the retro-transcribed viral genome into that of the host cell, to generate a provirus. The catalytic activities are catalysed by an enzyme that forms, at the extremities of the linear viral DNA, a tetrameric complex called stable synaptic complex (SSC). The enzyme catalyzes two successive reactions: the 3' processing reaction, which removes a GT dinucleotide from the 3' ends of the linear viral DNA and leaves reactive 3'OH ends, and the strand transfer reaction, during which the reactive 3'OH ends of viral DNA carry a nucleophilic attack to the phosphates of the target DNA allowing integration (Pauza et *al.*, 1990 ; Engelman et *al.,*1991). Integration occurs primarely into highly transcribed genes as the SSC associates with the cellular cofactor LEDGF/p75. Indeed, LEDGF/p75 acts as a targeting factor of the integration site (Ciuffi et *al.*, 2005) via its interaction with integrase, through the integrase binding domain (IBD) (Busschots et al., 2005), its domain of interaction with euchromatin (Gijsbers et *al.*, 2011) and its ability of targeting highly spliced transcriptionally active genes (Singh, P. K. et *al.*, 2015). Integrase is organised in three domains: the N-terminal domain (NTD, amino acids 1 to 46), the catalytic core domain (CCD, amino acids 56 to 186) and the C-terminal domain (CTD, amino acids 212 to 288) connected to each other through flexible linkers (Delelis et *al.*, 2008; Craigie and Bushman, 2012). The NTD adopts an helix-turn-helix fold and contains a HHCC motif that coordinates zinc cation, important for enzyme multimerization and structure stabilisation (Eijkelenboom *et al.,* 1997; Zheng et *al.*,1996). The CCD, which contains the catalytic triad D_{64}, D_{116}, and E_{152}, known as the $D,DX_{35}E$ motif, is responsible for catalysis, viral and cellular DNA binding and LEDGF binding (Heuer et *al.*,1997 ; Esposito & Craigie, 1998 ; Chen et *al.,* 2006, Busschots et al., 2007). It adopts an RNaseH fold and coordinates two Mg^{2+} in the active site (Dyda et *al.*, 1994; Bushman et *al.*, 1993)). Finally the CTD is involved in DNA binding (Engelman et *al.,* 1994; Lutzke et *al.*, 1994), multimérisation (Jenkins et al., 1996), interaction with reverse transcriptase (Wu et *al.*, 1999; Hehl et *al.*, 2004; Zhu et *al.*, 2004; Wilkinson et al., 2009) and with viral genomic RNA (Kessl et *al.*, 2016). This domain is the less conserved across different retroviruses and retrotransposon integrases (Cannon et *al.*, 1996; Kulkosky et *al.*, 1992; Johnson et *al.*, 1986) but, despite the degree of conservation of CTDs, the tree dimensional arrangement is similar: it adopts a Src homology 3 (SH3) fold (Eijkelenboom et *al.*, 1995).

While crystal structures of one or two domains are available (for review see Li et *al.*, 2011 and references therein; Chen et *al.*, 2000; Wang et *al.*, 2001), the complete structure of HIV integrase remains unsolved. Recently, though, two functional complexes between the wild type full-length integrase, the cellular cofactor LEDGF/p75 with (Maillot et *al.*, 2013) or without (Michel et *al.*, 2009) the integrase binding domain of the cellular cofactor INI1 (INI1-IDB) have been purified and analysed by mass spectrometry and cryolectromicroscopy (CryoEM). Based on this model, the orientation of the domains of HIV IN appears quite different from models previously proposed (Gao et *al.*, 2001; Karki et *al.*, 2004), raising the issue of which one most realistically captures the actual organisation of the protein. Indeed, the comparison of the crystal structures to the CryoEM model reveals a high flexibility of IN, mainly in the linkers between domains. More recently, another model was obtained by CryoEM analysis of complexes between an HIV-1 IN N-terminally extended, which facilitates its purification (Li et *al.,* 2014), and a branched DNA substrate mimicking the strand transfer complex (Passos et *al.*, 2017). This model is different from the two previous ones and could represent a successive catalytic step with respect to the other two complexes, further complicating the picture and emphasizing the need of additional caracterisation of this enzyme and the complexes it forms to achieve its tasks.

Since IN is expressed and assembled in virions as a 160 kDa Gag-Pol polyprotein precursor and catalyzes the integration of the proviral DNA as a 32 kDa mature protein, various mutations in IN have been shown to have pleiotropic effects. For this reason, mutations have been grouped in two classes: class I mutations are those which specifically affect the catalytic activity of IN; class II includes all the mutations affecting different stages of the HIV replicative cycle, such as maturation and

morphogenesis of viral particles (Engelman et *al.*, 1995; Bukovsky and Gottlinger, 1996; Quillent et *al.*, 1996; Kessl et *al.*, 2016), reverse transcription (Zhu et *al.*, 2004, Dobard et *al.*, 2007, Wilkinson et *al.*, 2009) and nuclear import of reverse transcribed products (Devroe et *al.*,2003 ; Zaitseva et *al.*, 2009 ; Jayappa et *al.*,2011).

Retention of functionality in proteins presenting sequence variability, as HIV proteins, relies on the existence of coevolution networks that allow counterbalancing the potential deleterious effect of one mutation by the introduction of one or more compensatory mutations in another position of the protein. The positions harbouring the two mutations are structurally and functionally related, providing information about the arrangement of the protein (Galli *et al.,* 2010 ; Woo et *al.*, 2014). Indeed, sequence variation due to misincorporation occurring during reverse transcription, to hypermutagenesis by cellular restriction factors, and to pervasive recombination occurring throughout the HIV genome constantly challenges retention of functionality (Preston et *al.*, 1988; Lecossier et *al.*, 2003; Hu et *al.*, 1990). Extensive coevolution of structurally and functionally related parts of the protein is the solution for the virus to conciliate genetic diversity and maintain of functionality. Recently, in the course of a study on the recombination in the Env gene between HIV-1 M primary isolates (Simon-Loriere et *al.*, 2009), we found that recombination breakpoints clustered in a particular way along the genome. Recombination in the IN portion of the pol gene were under represented, suggesting that a counterselection of the recombinants in that region had occurred probably because coevolution networks were broken.

In this present work, we exploited the natural genetic diversity of HIV to proceed to a functional characterisation of the CTD of HIV-1 integrase. Starting from the considerations made above, we have constructed chimerical IN between primary isolates from HIV-1 group M and group O and analysed their efficiency of reverse transcription and integration in cell infection tests in culture. The rationale was that, if coevolution networks were perturbed by virtue of the chimerical nature of the protein, a functional default would have been highlighted. We focus on the involvement in coevolution of the C-terminal domain, involved in the binding with viral and cellular DNA. This domain contains two regions fairly conserved in lentiviruses and essential for HIV-1 replication (^{235}WGPAKLLWKGEGAVV^{250} and ^{259}VVPRRK^{264}) (Cannon et *al.*, 1996). Sequence diversification in this domain can therefore have dramatic consequences on the global activity of the protein. We report here, that non-conserved residues in the CTD are essential for integration, as we highlighted a motif of four residues, which could be responsible for the recognition of the cellular DNA. The remarkable possibility of permutation of the residues in this motif without affecting IN functionality, underlines the importance of flexible regions, since this feature allows conciliating sequence diversification in HIV-1 and preservation of functionality.

Materials and Methods

Plasmids and molecular cloning of the parental strains

To study the functionality of HIV-1 IN enzyme, we used the pCMVΔR8.91 (Zuffery et *al.*, 1997) transcomplementation plasmid (referred herein as p8.91) modified as follows: we inserted the MluI and the BspEI restriction sites 18 nt downstream the beginning and 21 nt upstream the end of the RT coding sequence, respectively. This leads to three changes in amino acids in the RT (E_6T, T_7R and A_{554}S). Along with the SalI site, present downstream the end of the *pol* gene, these sites define two exchangeable cassettes: one encompassing the RT coding sequence (MluI-RT-BspEI, 1680 bp), thereafter called RT, and one encompassing the IN coding sequence (BspEI-IN-SalI, 1561 bp), thereafter called IN. The RTIN sequences of the two primary isolates used for this study were amplified and cloned into p8.91 between MluI and SalI sites to generate the corresponding parental transcomplementation plasmids. To facilitate the cloning of IN sequences (chimerical or mutant), we used the BspEI and SalI restriction sites.

Two plasmids were constructed for standard curves amplification in the different qPCR assays. One, called pJet-1LTR, for the detection of late and early product of reverse transcription was obtained by inserting the sequence of the LTR from the pSDY-dCK (see below) and the psi region in the plasmid pJet with the pJetPCRcloning kit (Thermo scientific). The second, pGenuine2LTR, possesses the perfect junction (unprocessed) U5-U3 and was constructed by Eurofins Genomics (Germany).

To produce the genomic RNA of the lentiviral vectors we modified a lentiviral vector previously described for the HIV–driven evolution of cellular genes (pSDY-dCK; Rossolillo et *al.*, 2012), in which the dCK gene was replaced by those coding for RPF, to monitor the efficiency of transfection by fluorescence microscopy, and the U3 sequence of HIV-1 in the 5'LTR in the plasmid was replaced by that of RSV (pSDY-RSV5-RFP, referred herein as pSDY).

Cells and viral strains

HEK-293T cells were obtained from the American Type Culture Collection (ATCC) and grown in Dulbecco's Modified Eagle's Medium (Gibco) supplemented with 10% foetal calf serum and 100 U/ml pennicillin-100 mg/ml streptomycin (Thermo Fisher) at 37°C in 5 % CO_2.

We used primary isolates from HIV-1 group M subtype A2 (GenBank accession # AF286237, named hereafter isolate A) and subtype C (GenBank accession #AF286224, named hereafter isolate C) obtained from the NIH AIDS Research and Reference Reagent Program. We also used a CRF02_AG (AAS638), a primary isolate from subtype B (AiHo, named hereafter isolate B) and the primary isolate RBF 206 (Genbank accession #KU168298, , named hereafter isolate O) from HIV-1 group O, obtained from J.C. Plantier, Virology Unit at CHU de Rouen associated to the French National HIV Reference Center.

Construction of IN chimeras

Chimerical integrases between primary isolates from HIV-1 group M subtype A2 and HIV-1 group O RBF 206 were constructed through overlapping PCR. Briefly, each chimerical gene is obtained from two fragments (5' fragment and 3' fragment of IN gene), of a given phylogenetic origin. Primers corresponding to the internal sequence (therefore called internal primers, forward and reverse) of the gene, overlapping the sequence where the phylogenetic switch is desired, are complementary and can allow the two fragments (5' and 3' fragments) to hybridize. These primers are used together with primers annealing at the extremities of the IN cassette, called external primers (forward and reverse), to amplify independently each fragment containing the overlapping sequence. These fragments (5' and 3') are then mixed in a subsequent PCR where primers are added only after few cycles. Full-length chimeric gene is reconstituted and amplified by adding only the external primers. The different full-length chimeric genes are digested with BspEI and SalI and cloned into the p8.91 plasmid.

A similar protocol has been used to produce single amino acid mutant integrases: the desired mutations were directly inserted in the sequence of the forward and reverse internal primers. The full-length mutated gene is obtained and cloned as indicated above.

Generation of pseudotyped viral particles

Pseudotyped lentiviral particles were produced by co-transfection of HEK 293T cells with p8.91 with two other transcomplementation plasmids, one coding for the genomic RNA (pSDY) and one coding for the envelope G glycoprotein of the vesicular stomatitis virus (pHCMV-G, Naldini et al., 1996), with the polyethylenimine method (PEI, MW 25000, linear; Polysciences, Warrington, PA, USA). HEK 293T were seeded at $5x10^6$ per 100-mm diameter dish and transfected 16-20 h later. The medium was replaced 6h after transfection, and the vector supernatants were recovered 48 or 72 h later, 0.45 µm filtered and the amount of p24 (CA) present in supernatants was quantified by ELISA (InnotestTM HIV Antigen mAB, International Genetic Technologies, Chilly-Mazarin, France). (Figure 1A)

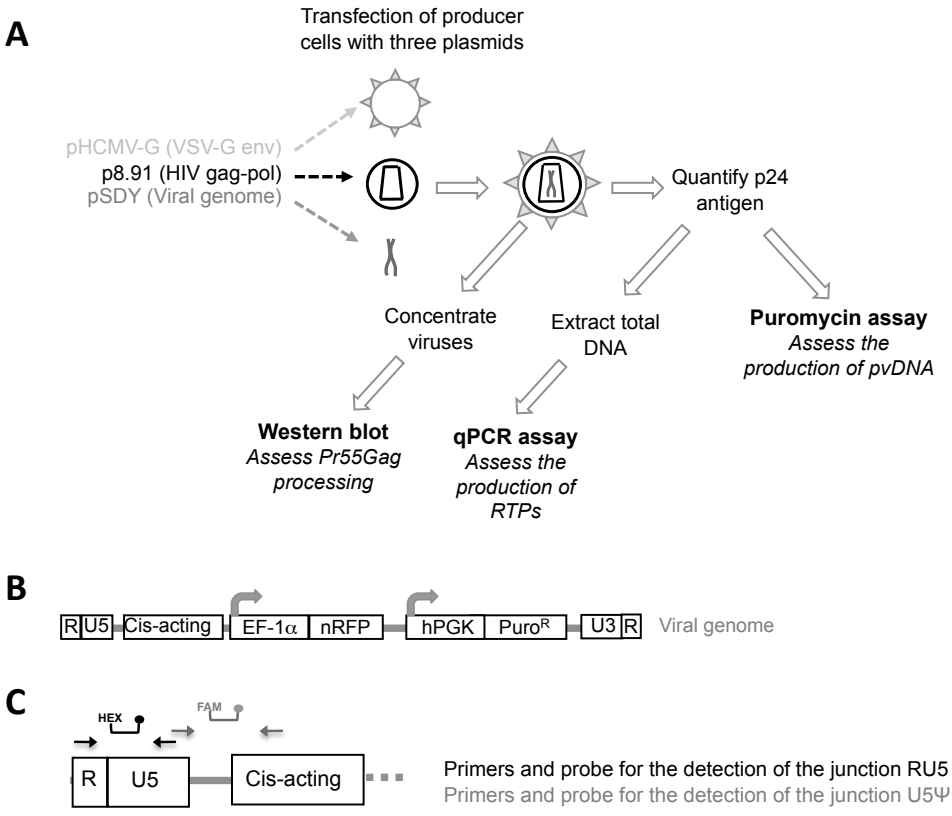


Figure 1: Panel A. Workflow used to study functionality of MO chimeras. **Panel B.** Viral genome, expressed by the pSDY. **Panel C.** Primers and probes for the detection of the viral LTR.

Western blotting

Western blot analysis was carried out on virions to assess the proteolysis of the Pr55Gag and Pr160GagPol precursors. 1.5 mL of viral supernatant was centrifuged through 20% sucrose, and the virions pellet was lysed in Laemmli buffer 1.5X. Viral proteins were separated on a CriterionTM TGX Strain-Free 4-15% gradient gel (Biorad) (TGS, 150V, 45 min), blotted on a PVDF membrane (TGS/Ethanol 10%, 200 mA, 1.5 h) and probed with a mouse monoclonal anti-CA antibody (NIH AIDS Reagent Program, #3537) to detect the viral capsid, the Pr55Gag unprocessed polyprotein and CA-containing proteolytic intermediates. An anti-mouse HRP-conjugated secondary antibody was used to probe the membrane previously incubated with anti-CA. Membranes were incubated with ECL reagent (Thermo Fisher) and WB were imaged on a Biorad Chemidoc Touch and analysed with the Biorad Image LabTM software.

Transduction and assessment of viral functions

Two functional assays were used, one to monitor the reverse transcription activity (qPCR assay), the other to quantify the level of proviral DNA (puromycin resistance assay / Alu PCR assay).

- *qPCR assay*

Non-internalised DNA was removed by treatment of the lentiviral vectors supernatants with 200U/ml of Benzonase® nuclease (Sigma-Aldrich) in the presence of 1 mM MgCl2 for 1 h at 37°C. The equivalent of 200 ng of p24 treated vectors were used to transduce 0.5 x 10^6 HEK 293T cells by spinoculation for 2 h at 32°C, 800 rcf, with 8µg/mL polybrene (Sigma-Aldrich). After 2 h the supernatant was removed, cells were resuspended in 2 mL of DMEM and plated in 6-well plates. After 30 h, cells were trypsinised

and pelleted. Total DNA was extracted with UltraClean® GelSpin® DNA Extraction Kit (Ozyme) adding a plasmid containing the GST sequence as internal control (pGEX). Two duplex qPCR assays were used, one to quantify early (detection of the RU5 junction) and late (detection of the U5Ψ junction) reverse transcription products (Figure 1C) and one to normalise for the quantity of cells employed in the assay (detection of β-actin DNA) and for fluctuations in the efficiency of the extraction step (detection of the exogenous GST sequence). The qPCR assays were designed with the Taqman® hydrolysis probe technology using the IDT Primers and Probes design software (International DNA technologies, Leuven, Belgium / PrimerQuest Tool), with dual quencher probes (one internal ZENTM quencher and one 3' Iowa BlackTM FQ quencher). All primers and probes were synthesised by IDT (Table 1). QPCRs were realised with the iTaq Universal Probes Supermix (Biorad) on a CFX96 (Biorad) thermal cycler with the following cycling conditions: initial Taq activation 3', 95°C followed by [denaturation 10", 95°C; elongation 20", 55°C] x 40 cycles. Standard curves and analysis were carried out with the CFX Manager (Biorad). The early and late reverse transcription products quantities were normalised with the two control targets, β-actin and GST. Copy numbers of DNA were determined in reference to a standard curve prepared by serial dilutions of the corresponding plasmid (pJet-1LTR and pGEX).

duplex	target	primer/probe	sequence (5'-3')	fluorophore
duplex I (quantification)	U5Ψ	U5Ψ-F*	GTGACTCTGGTAACTAGAGA	-
	U5Ψ	U5Ψ-probe	CGCTTTCAAGTCCCTGTTCGGG	FAM
	U5Ψ	U5Ψ-R**	GAGAGCTCCTCTGGTTTC	-
	RU5	RU5-F	CAGATCTGAGCCTGGGAG	-
	RU5	RU5-probe	AAGCAGTGGGTTCCCTAGTTAGCC	HEX
	RU5	RU5-R	GGCACACACTACTTGAAGC	-
duplex II (normalisation)	GST	GST-F	CGTTATATAGCTGACAAGCACAAC	-
	GST	GST-probe	AGAGCGTGCAGAGATTTCAATGCTTG	FAM
	GST	GST-R	GCAATTCTCGAAACACCGTATC	-
	ACTB	IDT pre-designed assay, Hs.PT.56a.40703009.g/exon 6		HEX

* forward
** reverse

Table 1: Primers and probes used in the qPCR assay.

- *Puromycin assay*

0.5 x 10^6 HEK 293T cells were transduced with a volume of lentiviral vectors corresponding to 0.2 ng of p24 by spinoculation 2h at 32°C, 800 rcf, with 8µg/mL polybrene (Sigma-Aldrich). After 2 h the supernatant was removed, cells were resuspended in 7 mL of DMEM and plated in 100 mm diameter dishes. After 30 h, puromycin was added at a final concentration of 0.6 µg/mL, clones were allowed to grow for 10 to 12 days and then counted.

- *Alu PCR assay*

Total DNA extracted from cells transduced for the qPCR assay was used for the Alu PCR assay, as already described (Vozzolo et *al.*, 2010). Briefly, two rounds of amplification were used. The first round of PCR, with Alu-forward primer and Psi reverse primer, to amplified Alu-LTR fragments (Table 2). The cycle parameters were as follows: 95°C for 3min, [95°C for 30s, 55°C for 30s, 72°C for 3min30s] x15, 72°C for 7min. Samples were diluted to 1:10 and 2µL were used in the second qPCR, to detect the viral LTR, as describe above for the detection of the RU5 junction.

stage	target	primers/probes	sequence (5'-3')	fluorophore
1st PCR	ALU-LTR	Alu forward	TGCTGGGATTACAGGCGTGAG	-
	ALU-LTR	Ψ reverse	GCTCCTCTGGTTTCCCTTTC	-
2nd qPCR	RU5	RU5-forward	see above, table 1	
	RU5	RU5-probe		
	RU5	RU5-reverse		

Table 2: Primers used for the Alu PCR assay.

Quantification of 2-LTR circles

Non-internalised DNA was removed by treatment of Benzonase® nuclease as for the qPCR assay (see above) and 0.5 x 10^6 HEK 293T cells were transduced with a volume of lentiviral vectors corresponding to 1μg of p24 by spinoculation as described above. After 30 h, cells were trypsinised and pelleted. Total DNA was extracted with UltraClean® GelSpin® DNA Extraction Kit (Ozyme). The late reverse transcription products (detection of the U5Ψ junction) was assessed as for the qPCR assay (see above) and two qPCR assays were used, as previously described, to quantify the total amount of 2 LTR circles and the quantity of 2 LTR circles with a perfect palindromic junction, with a primer overlapping the 2LTRc junction. The qPCR assays were designed with the Taqman® hydrolysis probe technology using the IDT Primers and Probes design software. All primers and probes were synthesised by IDT (Table 3). QPCRs were realised with the iTaq Universal Probes Supermix (Biorad) on a CFX96 (Biorad) thermal cycler with the following cycling conditions: initial Taq activation 3'/95°C – [denaturation 10"/95°C, elongation 20"/55°C] x40. Standard curves and analysis were carried out with the CFX Manager (Biorad). Copy numbers of the different forms of viral DNA were determined with respect to a standard curve prepared by serial dilutions of the corresponding plasmid (pGenuine2LTR).

target	primers/probes	sequence (5'-3')	fluorophore
2LTRc	2LTR forward	CCCTTTTAGTCAGTGTGGAA	-
2LTRc	2LTR probe	TTCACTCCCAACGAAGACAAGATATCCTT	FAM
2LTRc	2LTR reverse	GTAGCCTTGTGTGTGGTAGA	-
PJ	2LTR PJ forward	TGTGGAAAAATCTCTAGCAGTAC	-
PJ	2LTR probe	TTCACTCCCAACGAAGACAAGATATCCTT	FAM
PJ	2LTR reverse	GTAGCCTTGTGTGTGGTAGA	-

Table 3: Primers and probes used for the detection of total 2LTRc and 2LTRc with perfect palindromic junction (PJ).

Sequence alignments

We used 3366 sequences for alignment. The original HIV-1 group M sequences were downloaded from the Los Alamos National Laboratory (LANL) database. They are derived from different subtypes: A (249 sequences), B (2450 sequences), C (450 sequences), D (121 sequences), G (80 sequences), H (8 sequences), J (6 sequences), K (2 sequences). We aligned 49 sequences from HIV-1 group O. We downloaded 26 of them from the LANL database and the 21 other were obtained through the collaboration with J.C. Plantier, Virology Unit at CHU de Rouen associated to the French National HIV Reference Center. Sequence alignments were performed with the CLC sequence viewer 6 software. The alignment of the CTD of HIV-1 IN group M was uploaded in the WebLogo software, (http://weblogo.berkeley.edu/logo.cgi) to obtain the sequence logo of the positions 222, 240, 254 and 273 in the group M.

Statistical tests

All statistical analysis were performed on at least three independent experiments (transfection and transduction). The statistical analysis was carried out using the Prism 6 software (GraphPad). The values obtained for the chimeras were normalized using the values obtained for parental INA. Student test at one sample was used to evaluate whether the normalized mean values obtained with the chimeric and mutant integrases were significantly different from that obtained with the parental strain.

Assessment of the infectivity of *replication-competent viruses*

PNL4.3 plasmid (NIH AIDS Reagent Program, #114) were modified by replacing the sequence of the CTD of IN by those from the isolate A or the mutant of NKNK motif (KQKQ, KQNK, NQKK).

Replication-competent viruses were produced by transfection of HEK 293T cells with PNL4.3 plasmids with the polyethylenimine method (PEI, MW 25000, linear; Polysciences, Warrington, PA, USA). The medium was replaced 6h after transfection, and the vector supernatants were recovered 48 or 72 h later and used to infect cells (TZM-bL or CEM-SS).

- Detection of virus replication on TZM-bL cells

25 µl of TZM-bL cells (a HeLa cell clone genetically engineered to expresses CD4, CXCR4, and CCR5 and containing Tat-responsive reporter genes for firefly luciferase under regulatory control of an HIV-1 long terminal repeat) at 4x 10^5 cells /ml were infected with 25 µl of virus dilution as indicated. 50 µl of culture medium (DMEM 10% SVF) was added. After 48h, virus replication was detected by measuring Luc reporter gene expression. Briefly, 75µl of culture medium was removed from each well and 50 µl of Bright Glo reagent was added to the cells. After a 2-min incubation at room temperature to allow cell lysis, 100µl of cell lysate was transferred to 96-well black solid plates for measurements of luminescence (RLU) using a luminometer (Sarzotti-Kelsoe et *al.*, 2014).

- Detection of virus replication on CEM-SS cells

0.5 x 10^6 CEM-SS cells/5ml were infected with 1/25 virus dilution. After 5 days of culture, the percentage of infected cells were detected by intracellular p24 staining and flow cytometry analysis as previously described (Lederle et *al.*, 2014).

Results

Analysis of intergroup M/O chimeras

The functionality of chimerical integrases was evaluated as shown in Figure 1A. Conditional replication-defective VSV-pseudotyped HIV-1 derived vectors were produced by triple transfection with pHCMV-G (encoding the VSV-G envelope protein), pSDY (that leads to synthesis of the genomic RNA), and p8.91 (encoding the enzymatic and structural viral proteins, apart from the envelope). This last plasmid is the one encoding for the integrase and contains a cassette that was used to replace the different sequences of the integrases studied. The viruses are used to transduce HEK 293T cells in culture. Since the genomic RNA is deleted of all the genes encoding for viral proteins (Figure 1B), infection will be blocked after integration of the proviral DNA, as new particle cannot be formed. The genomic RNA contains the internal promoter for the human phosphoglycerate kinase (hPGK) that drives the expression of the gene for the puromycin N-acetyl-transferase that confers resistance to puromycin. The rationale for the assay is that, once reverse transcription is completed, if integration occurs, the proviral DNA will stably lead to the production of the puromycin N-acetyl-transferase and the individual cells can expand in the presence of puromycin, leading for each HEK 293T successfully transduced to the generation of one clone. Transient expression of the gene from non-integrated forms of the DNA reverse transcribed, in contrast, will not lead to the generation of clones in long-term cultures in presence of puromycin. Since mutations in the integrase can also affect reverse transcription and the number of integrated proviruses that can be generated is also dependent on the amount of pre-proviral DNAs available after reverse transcription, we express the efficiency of integration after normalization for the quantity of late products of pre-proviral DNA, estimated by qPCR as described in Methods.

Performing the test with four variants of p8.91 has tested the reliability of this approach. These variants carry either the RT and IN sequences of the parental primary isolate from subtype A of group M ($RTA^{(+)}$ $INA^{(+)}$ in Table 4), or the primary isolate RBF 206 from group O ($RTO^{(+)}$ $INO^{(+)}$), or the sequence coding for the catalytically inactive double mutant RTA, $D_{110}N$-$D_{185}N$ (Larder et *al.*, 1987) ($RTA^{(-)}$ $INA^{(+)}$) or for the catalytically inactive integrase mutant, $D_{116}A$ (Cannon et *al.*, 1996) ($RTA^{(+)}$ $INA^{(-)}$). As shown in Table 4 all the samples gave the expected results in qPCR and in the puromycin-

resistance assay, based on the properties of their RT and IN. In parallel, to further corroborate the results obtained with the puromycin-resistance assay, an Alu PCR assay was performed (Vozzolo et *al.*, 2010), showing a perfect consistence between the two assays (Table 4).

	RT (QPCR, % of HXB2)	**IN** ($puro^R$ clones, % of HXB2)	**IN** (Alu PCR, % of HXB2)
$RTA^{(+)}\,INA^{(+)}$	90 ± 21	60 ± 11	69 ± 14
$RTO^{(+)}\,INO^{(+)}$	51 ± 19	42 ± 16	39 ± 9
$RTA^{(+)}\,INA^{(-)}$	71 ± 37	0 ± 0	0 ± 1
$RTA^{(-)}\,INA^{(+)}$	0 ± 0	n.a.	n.a.

Table 4: Validation of the experimental system. Activity of RT (detected by qPCR) and IN (detected by the puromycin resistance test or by the Alu PCR assay) for the two isolates (A and O), and for catalytic mutants of RT (double mutant $D_{110}N$-$D_{185}N$) and IN (mutant $D_{116}A$) from isolate A.

In parallel, since a defect in processing of the Gag and Gag-Pol polyprotein precursors would result in impaired reverse transcription and integration, Western blot on the viral particles was also performed to monitor the degree of processing of the precursor Gag.

Eleven chimeras between primary isolates A and O were studied. The positions of the breakpoint between the two isolates spanned from amino acid 1 to 285 (Figure 2A), and the chimeras are named hereafter by the position of the breakpoint. Pr55Gag processing was comparable to that of the wt proteins for all chimeras starting from breakpoint position 106 and proceeding to the C-ter of the protein (Figure 2B and 2C). For chimeras with a breakpoint between the N-ter and position 71, processing was reduced to approximately 20 % that of the wt proteins. Expectedly, reverse transcription was strongly reduced and integration was almost undetectable in these chimeras (Figure 2D and 2F, respectively). For the chimeras from 106 to 285, depending on the chimeras, the results obtained for reverse transcription and integration differed. A progressive increase in the amount of reverse transcription products was observed moving toward the C-ter of the protein, reaching levels around 80% of those of wt integrase A with chimeras 212 and 285. Interestingly, for chimera 272, instead, reverse transcription was reduced to 30% approximately. No significant difference between the quantities of early and late reverse transcription products was observed for these chimeras, except for 272 (Figure 2E). For this chimera there is a significant difference between the quantities of early (56%) and late (32%) products, indicating that reverse transcription events are not expected to occur due to a probable defect in the strand transfer or in the degradation of the RNA template by RNaseH. The amounts of similar early and late products but lower than the parental level for chimeras 106 to 195 suggest that the lack of activity relates to a step prior to the first strand transfer. For integration, the differences among the chimeras were more marked. No integration products were observed for chimeras 106 and 137 while the level of integration was comparable to that of the wt proteins for the other chimeras, except for chimera 272 for which the events of integration was reduced to around 30% (Figure 2F).

To characterise the defects in integration observed for certain chimeras, we then focused on the chimeras that displayed a correct level of processing of Pr55Gag (chimeras 106, 137 and 272). Among these, we chose chimera 272 for a further characterization, because of the limited difference in the number of amino acid that differentiates this chimera and the almost fully functional chimeras 212 and 285. This choice also allowed a thorough characterisation of the whole CTD, the less studied domain of HIV-1 IN.

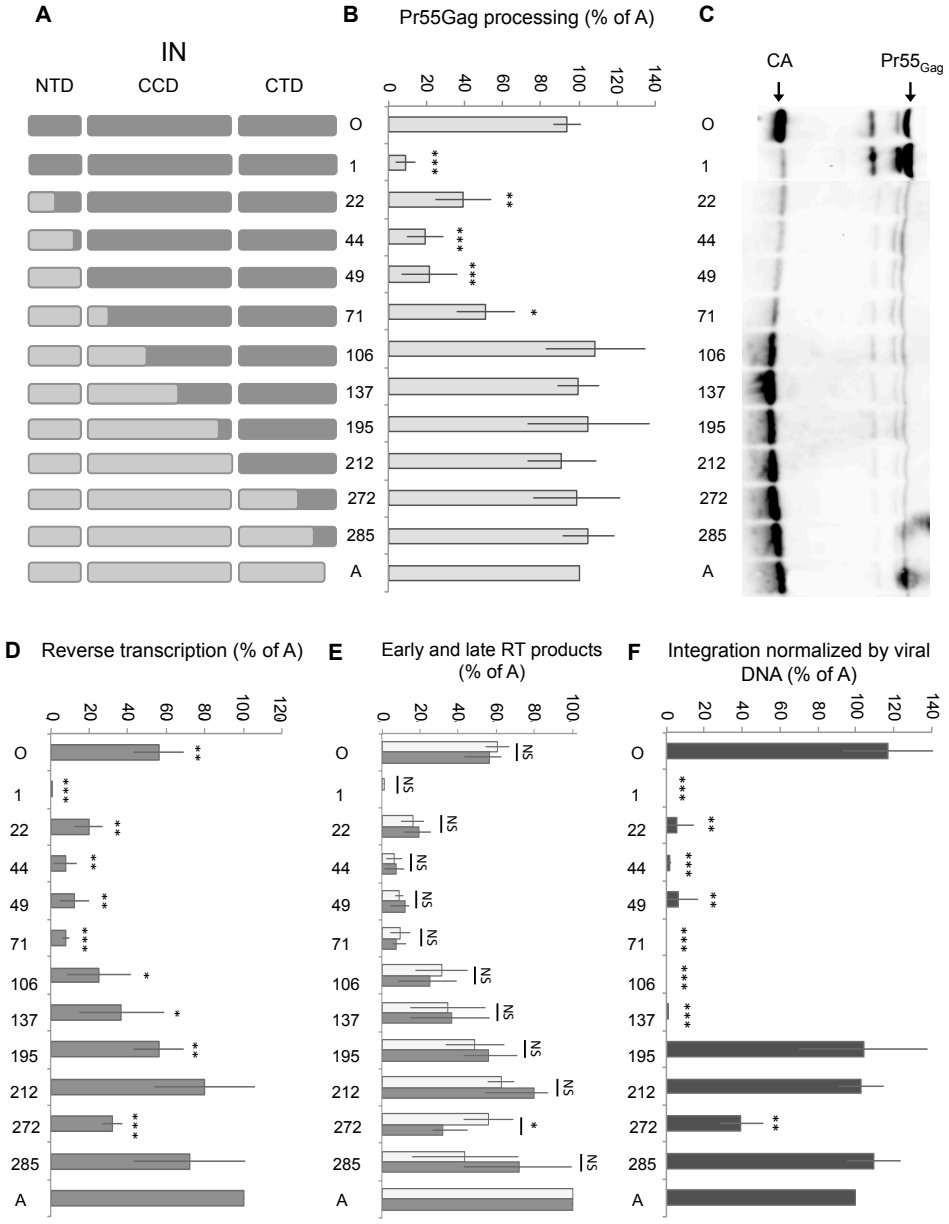


Figure 2: Functionality of viruses harbouring chimerical integrases. Panel A. Representation of the chimerical IN studied. The position of the breakpoint is given on the left, in aa from the beginning of the IN coding region. IN A is 288 aa long, while IN O is 298 aa. **Panel B.** Gag processing efficiency, estimated by the amount of CA compared to the amount of Gag precursors detected by Western blot. **Panel C.** Efficiency of reverse transcription (detection of the junction U5Ψ) for the different chimerical integrases. **Panel D**. Relative quantities of early (detection of the junction RU5) and late RT (detection of the junction U5Ψ) products. **Panel E.** Relative efficiency of integration normalized by the amount of total viral DNA, expressed as a function of A, fixed at 100%. Errors bars are standard deviations. Experiments have been repeated at least 3 times. Stars represent Student statistics.
(NS p value>5%, * p value<5%; ** p value<1%; *** p value<0,1%).

R	215	K
T	218	L
N	222	K
K	240	Q
N	254	K
G	255	S
K	273	Q
D	278	T
V	281	M
G	283	S

Table 5: Residues that differ between IN A and O from position 212 to 285. **Left raw.** Amino acid present in the IN from isolate A. **Middle raw.** Position of the residue on IN. **Right raw.** Amino acid present in the IN from isolate O.

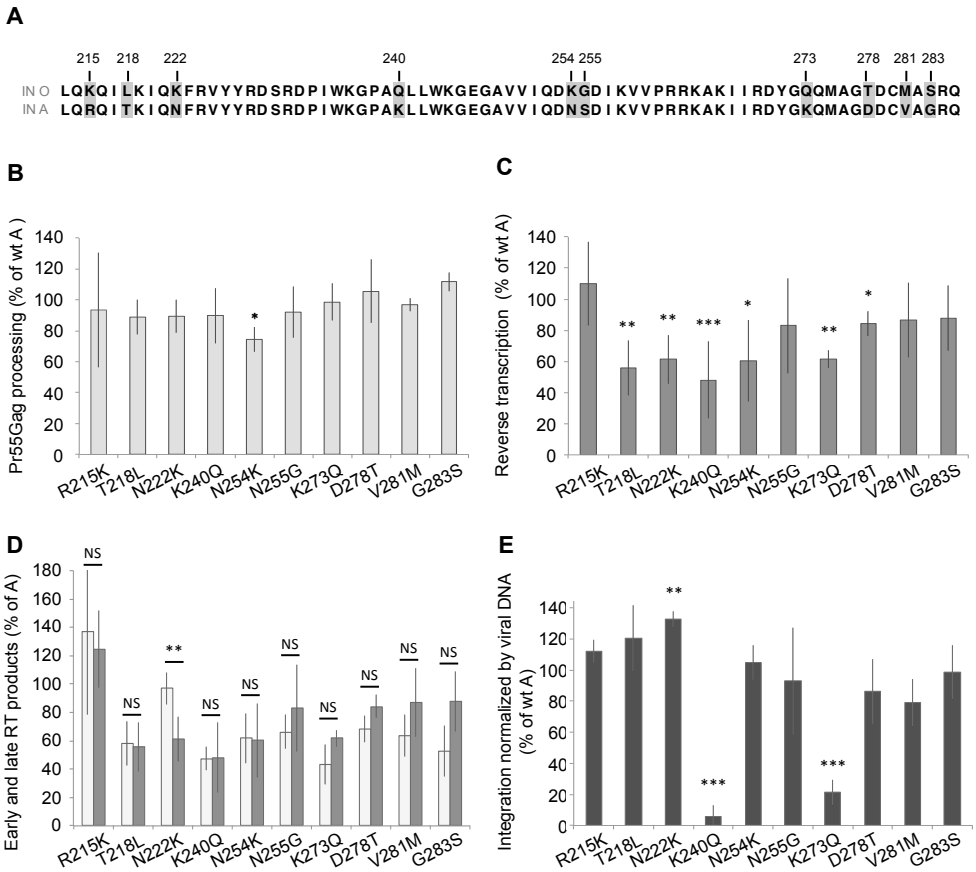


Figure 3: Functionality of viruses harbouring mutant integrases. Panel A. Sequence alignment of IN, sequence portion spanning amino acids 212 to 285, from isolates A and O. **Panel B.** Gag processing efficiency, estimated by the amount of CA compared to the amount of Gag precursors detected by Western blot. **Panel C.** Efficiency of reverse transcription (detection of the junction U5Ψ) for the different mutant integrases, named as the residue in IN A, its position and the residue present in IN O. **Panel D.** Relative quantities of early (detection of the junction RU5) and late RT (detection of the junction U5Ψ) products. **Panel E.** Relative efficiency of integration normalized by the amount of total viral DNA, expressed as a function of A, fixed at 100%. Errors bars are standard deviations. Experiments have been repeated at least 3 times. Stars represent Student statistics.
(NS p value>5%, * p value<5%; ** p value<1%; *** p value<0,1%).

Characterisation of the coevolution within the CTD

The isolates A and O used to generate the chimeras of Figure 2 differ, between positions 212 and 285, for 10 residues (Figure 3A). Each of these residues from isolate O was individually replaced in wt IN A and the ten point mutants tested, as described above, for processing of Pr55Gag, reverse transcription (early and late RT products) and generation of integrated proviruses.

All mutants presented comparable levels of processing of Pr55Gag (Figure 3B) and moderate decreases in reverse transcription that remained in the range of 90-50 % with respect to wt A IN (Figure 3C). The comparison of early and late RT products reveals no significant difference except for the mutant $N_{222}K$, which presents more early than late RT products (Figure 3D). Integration, instead, showed more contrasted results, being markedly decreased for two mutants (around 5 and 20 % for $K_{240}Q$ and $K_{273}Q$, respectively) while for all other mutants it ranged between 80 and 130 % that of wt integrase A (Figure 3F). In both mutants with the marked defect in functionality, the polymorphism introduced was the replacement of a K in isolate A by a polar residue (Q in both cases) that is present in isolate O at the corresponding positions. In isolate O, two K are present in positions where a polar amino acid (N in both cases) is present in isolate A (positions 222 and 254, Table 5). We therefore reasoned that the two K present at positions 222 and 254 of isolate O could exert the same function of the two K at positions 240 and 273 in isolate A.

To test this hypothesis, we generated the mutant of isolate A $N_{222}K/K_{240}Q/N_{254}K/K_{273}Q$. Integration with this mutant was comparable to what observed with wt A integrase (89 %) suggesting that the presence of the two lysines is required but that their positions can be changed within the motif constituted by positions 222/240/254/273. To test whether the presence of two K at any of the four positions of the motif is sufficient to ensure integration, we generated: a mutant with no K in the motif (NQNQ), the four possible mutants containing only one K at any of the positions of the motif, all the possible mutants with two or three K at each of the four positions of the motif and, finally, the mutant with four K, one in each of the four positions of the motif. The absence of K totally abolished integration, and the presence of a single K led on average to 19% of integration with respect to the wt integrase A (Figure 4A). In contrast, the presence of two or more residues led to levels of integration spanning from 75 to 137% those of wt integrase A.

When looking at the individual mutants within each class, significant variations were however observed (Figure 4B). Namely, while the mutants containing three K led to results not significantly different among the various samples, large variations were observed within the classes of mutants with two or one K. In the class of the mutants with a single K, the presence of the K in the third and fourth positions of the motif (positions 254 and 273) resulted in almost undetectable levels of integration, in contrast to what observed for the presence of a lysine at the first and second position of the motif (positions 222 or 240), suggesting that the presence of the K is less important in the third and fourth than in the first and second positions.

Consistent with this view, among the mutants with two K, the protein with K in these two positions (NQKK) was the one with the most dramatic reduction in integration (22% that of the parental integrase A). The only other mutant with two K that displayed a significant reduction in integration was mutant NKKQ, with a level of integration of 48%. All the other mutants with two K led to an average integration of 110% that of wt integrase A.

Overall, these results show that the presence of two K in the motif is required for integration by the integrase A and, even if it is possible to permute the positions of the K retaining integration in several cases, some positions have a significantly different impact on the integration process.

K240 and K273 are required in group M isolates

Alignment of sequences of group M integrases reveals the strong conservation of the NKNK motif across group M (Figure 5A). To broaden to other integrases of group M the validity of the observation made with integrase A we replaced the NKNK motif in the integrases from primary isolates of different group M subtypes (subtype C, subtype B, CRF 02) by the motif deprived of K described above (NQNQ). In all cases a dramatic drop in integration was observed with respect to the corresponding wt integrase, confirming the results obtained with isolate A (Figure 5B). Integration was almost undetectable with the mutated integrases B and C and reduced to 28% for integrase CRF02. The importance of the two K of the motif is therefore confirmed in all phylogenetic groups the most widespread in the epidemics (A, B, C, and CRF 02 being responsible for 77% of the HIV infections worldwide) (Buonaguro et *al*., 2007).

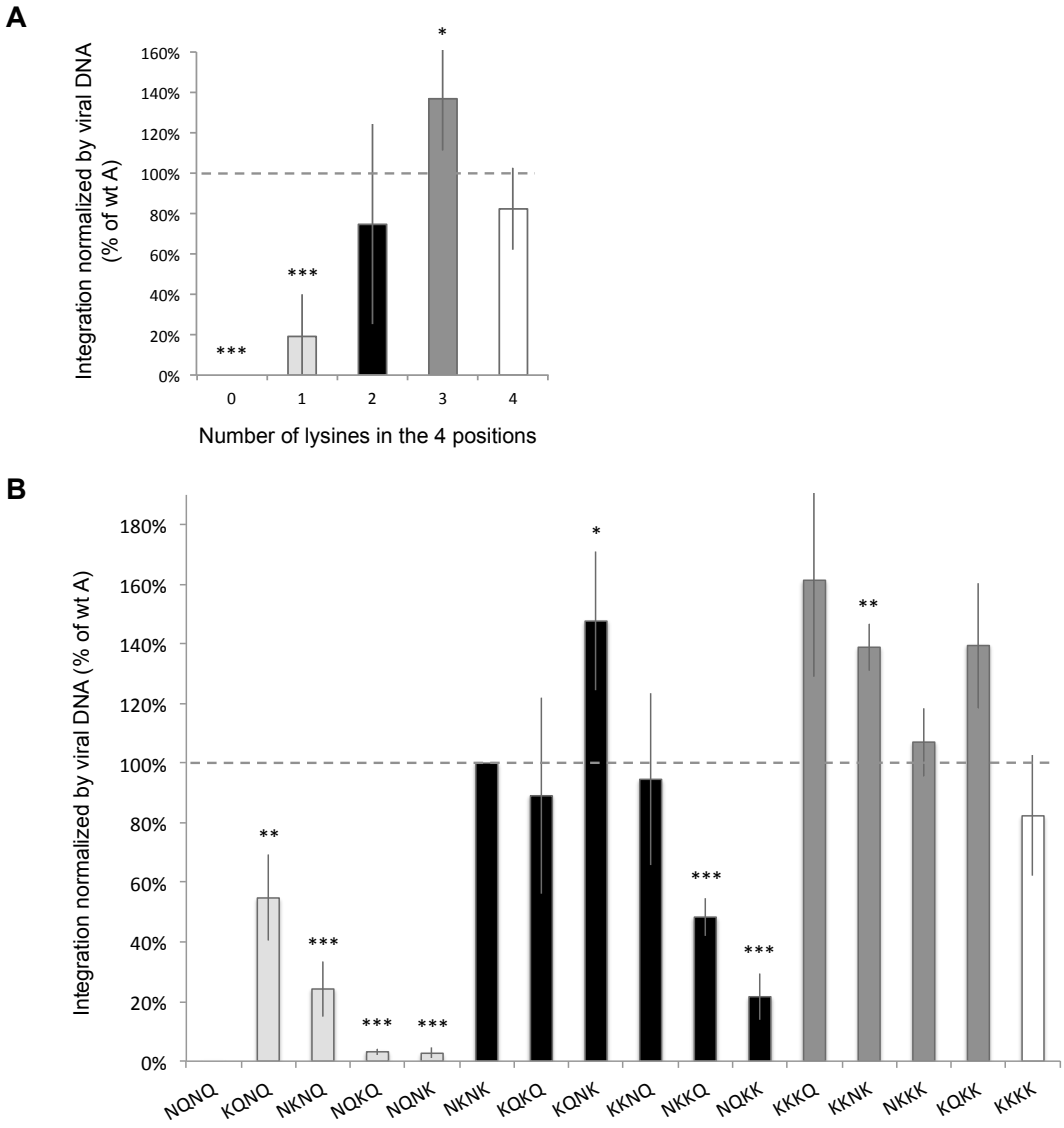


Figure 4: Panel A. Relative efficiency of integration, normalized by the amount of viral DNA, for the IN mutants grouped by the number of lysine present in the positions 222, 240, 254, 273. **Panel B.** Relative efficiency of integration, normalized by the amount of viral DNA, for the integrase mutants which presents 0, 1, 2, 3 or 4 K in the positions 222, 240, 254, 273. Errors bars are standard deviations. Experiments have been repeated at least 4 times. Stars represent Student statistics (NS p value>5%, * p value<5%; ** p value<1%; *** p value<0,1%).

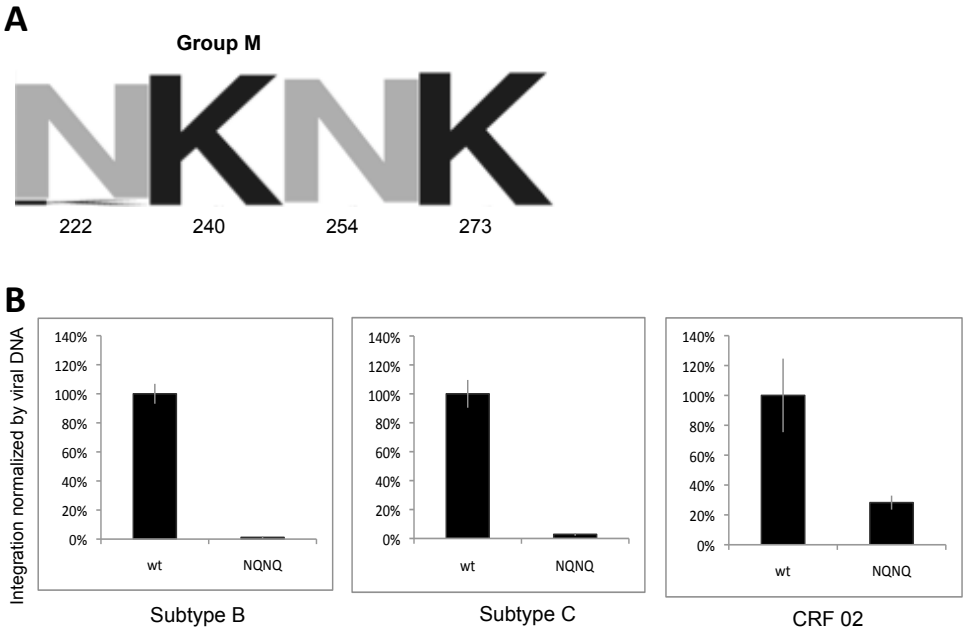


Figure 5: Panel A. Conservation logo done with the WebLogo software, http://weblogo.berkeley.edu/logo.cgi. The sequence at the positions 222, 240, 254 and 273 in group M integrases (3390 sequences from LANL database) have been aligned with CLC sequence viewer. **Panel B.** Relative efficiency of integration normalized by the amount of viral DNA of the mutant 0K (NQNQ) compared to the wt integrase of different group M isolates (subtype B, subtype C, CRF02). The graphic are disposed according to the phylogenetic origin of the primary isolate used, indicated below. Errors bars are standard deviations. Experiments have been repeated at least 3 times.

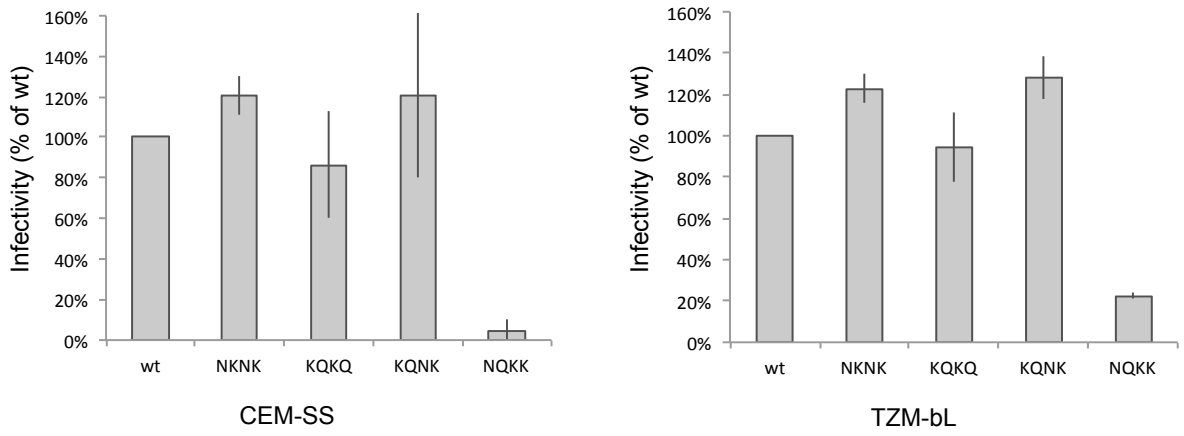


Figure 6: Infectivity of several mutations of the motif, cloned in the replicative strain PNL4-3, expressed as a percentage of the wt. **Left panel.** Replication-competent viruses was used to infect CEM-SS cells. The infectivity represents the percentage of infected cells compared to wt, detected by intracellular p24 staining and flow cytometry analysis . **Right panel.** Replication-competent viruses was used to infect TZM-bL cells. The infectivity represents the virus replication compared to wt, detected by measuring Luc reporter gene expression. The error bars represent the standard deviations of sample duplicates.

The NKNK motif in replication-competent viruses

The sequences encoding for some of the functional mutants (Figure 4B) were then inserted in replication-competent pNL4.3 viruses and infecting CEM-SS or TZM-bL cells in culture tested infectivity of the chimerical viruses. Infection was monitored as described in Materials and Methods. We focused on mutants that contain two K in ectopic positions with respect to the canonical motif of group M, NKNK. The mutant markedly impaired in integration despite the presence of two K in the motif (NQKK), and two mutants that are still integration-competent (mutants KQNK and KQKQ) were inserted in pNL4.3 viruses. A variant of pNL4.3 carrying the CTD sequence of wt integrase A (NKNK) was also constructed and used as a control in parallel to the use of wt pNL4.3. The results obtained confirmed the observations made in the single infection cycle, with all the viruses that displayed levels of infectivity comparable to those of wt pNL4.3 viruses except NQKK that had a reduction of infectivity reduced to approximately 20% (Figure 6). The experiments for each cell lines have only been done, for now, in duplicates, so we cannot conclude on the significance of the results, even if they are quiet similar to those obtained in a single infection cycle.

Characterisation of the motif NKNK

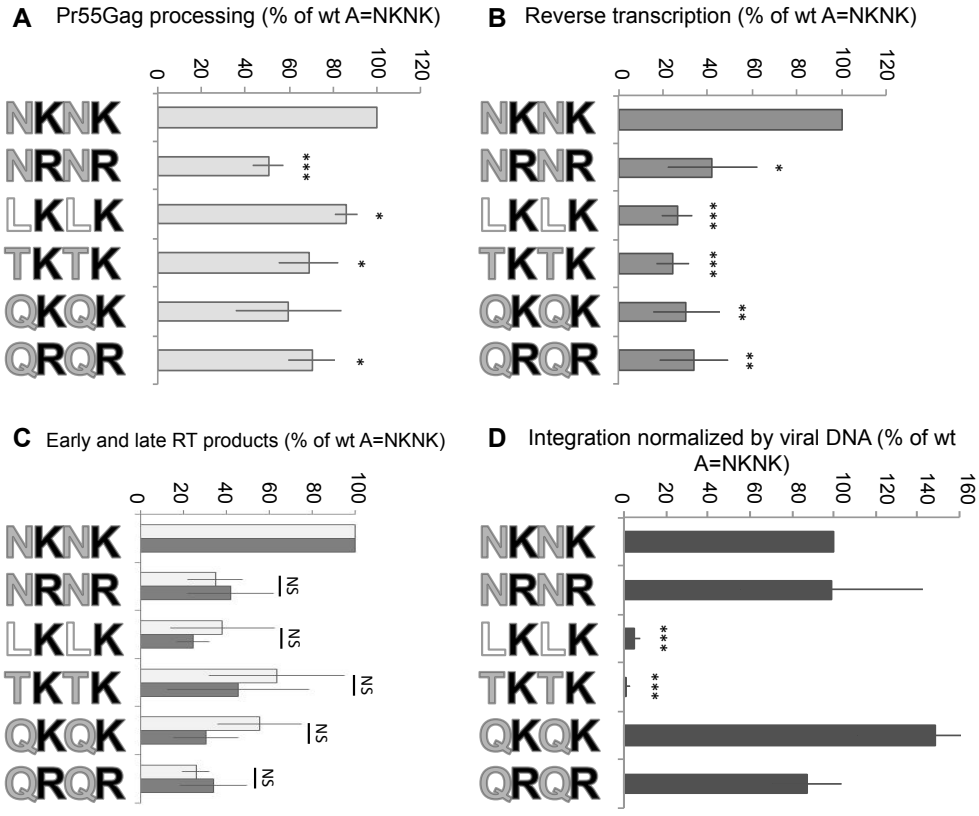


Figure 7: Panel A. Gag processing efficiency, estimated by the amount of CA compared to the amount of Gag precursors detected by Western blot, expressed as the function of the parent A, NKNK. The residue present at the four positions of the motif for each sample are indicated to the left. **Panel B**. Efficiency of reverse transcription (detection of the junction U5Ψ) for the different mutant integrases, expressed as a function of NKNK, fixed at 100%. **Panel C.** Relative quantities of early (detection of the junction RU5) and late RT detection of the junction U5Ψ) products, expressed as the function of NKNK, fixed at 100%. **Panel D.** Relative efficiency of integration normalized by the amount of total viral DNA, expressed as a function of NKNK, fixed at 100%. Errors bars are standard deviations. Experiments have been repeated at least 3 times. Stars represent Student statistics.
(NS p value>5%, * p value<5%; ** p value<1%; *** p value<0,1%).

The specificity of the amino acids that constitute the NKNK motif in group M integrases was then investigated. To understand if another positive amino acid can substitute the lysine residues present at positions 240 and 273, we have replaced the two K by two arginine residues (NRNR mutant). The mutant was as functional as wt integrase A (Figure 7F) underscoring the importance of the positive charge at positions 240 and 273. To test the importance of the other components of the motif, we replaced the two N residues at positions 222 and 254 by amino acids of comparable size that either retain polarity albeit with a different polar group as threonine (mutant TKTK) or that abolish the polar nature of N, as leucine (mutant LKLK). In both cases, integration was almost abolished (Figure 7F) indicating that not only the presence of a polar amino acid but also the nature of the polar group present at positions 222 and 254 is as important as that of the positively charged K at positions 240 and 273. The need for a polar residue carrying the NH2 group at positions 222 and 254 was then further confirmed by the replacement of the two N by two Q (mutant QKQK) that led to levels of integration comparable to those of the wt integrase. Finally, the quadruple mutant QRQR that displayed levels of integration comparable to the wt integrase despite the replacement of both, the K and the N of the original motif by R and Q, respectively, has further supported these conclusions (Figure 7F).

Mutations in the integrase are known to have frequently pleiotropic effects. However, the alteration of the polar and basic nature of the residues in the motif seems to impact mostly integration since Pr55Gag processing and reverse transcription gave similar results irrespective of whether the mutations introduced conservation the basic, or polar, nature of the residues (Figures 7B, 7C and 7D). However, all mutants displayed levels of processing and of reverse transcription lower than wt integrase A indicating that, expectedly, the natural sequence motif NKNK is likely the most efficient for the infectious process.

Identification of the role of the motif NKNK in integration

To identify which step of the integration process is affected by the mutations in the NKNK motif, for some samples we determined the amount of two LTR circles (2LTRc) and analysed the nature of the LTR-LTR junction in these circles. The first test measures the amount of DNA that has been imported into the nucleus but has not been integrated. In integration-defective mutants, an increase in this amount is therefore indicative of a default in integration *per se* but of a correct nuclear import process. In contrast wt or lower than wt levels of 2LTR circles for integration-defective mutants suggest an impaired process of nuclear import. The second analysis (the study of the nature of the LTR-LTR junction in 2LTRc) is informative of the efficiency of 3' processing. 2LTRc can indeed be generated starting from both unprocessed and processed 3' ends. In the first case, the 2LTRc will contain perfect palindromic junctions while in the second they will present imperfect palindromic junctions. A high proportion of 2LTRc with perfect palindromic junctions is therefore indicative of an inefficient 3' processing.

These analyses were carried out on five mutants of the NKNK motif, one mutant of the most important K residue of the motif (the one at position 240) that retained the integration competence (KQNK), one that lost it (NQNK), the integration-deficient double mutant deprived of lysines (NQNQ), and two integration-deficient mutants of the polar residues of the motif (TKTK and LKLK). The two parental integrases A and O and the catalytic integration-deficient mutant D116A were used as controls.

The two mutants of the polar residues both showed an accumulation of 2LTRc to levels significantly higher than for the parental integrases (A and O) and closer to those found for the nuclear import-competent but integration-deficient mutant $D_{116}A$ (Figure 8A). This suggests that nuclear import is not severely impaired and that other steps must be deficient in these mutants to account for the dramatic default in integration observed. The two integration-deficient mutants NQNK and NQNQ, instead, displayed levels of 2LTRc only slightly higher than those observed for the parental enzymes. Being integration deficient mutants, this probably reflects the existence of a defect at the level of nuclear import.

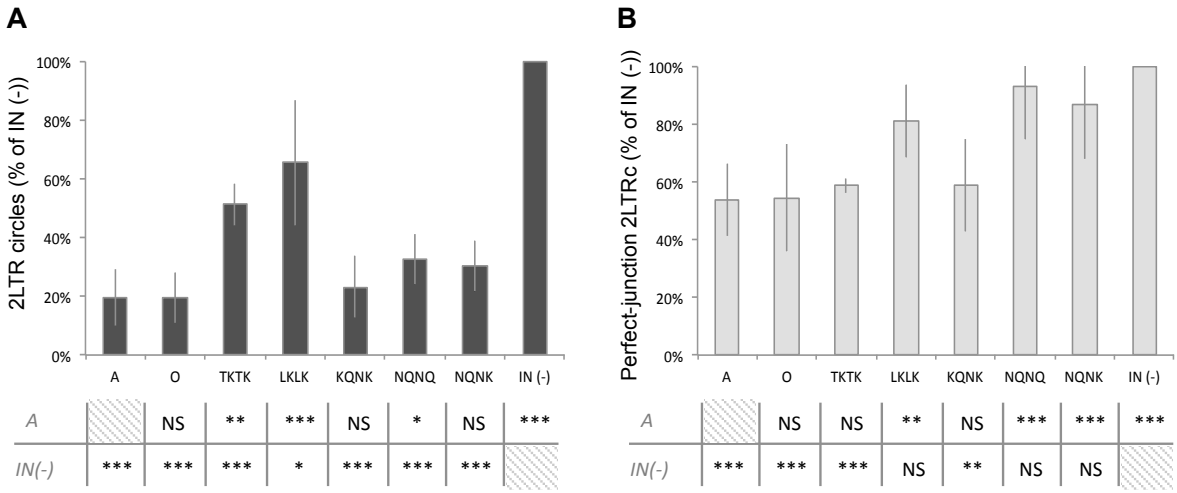
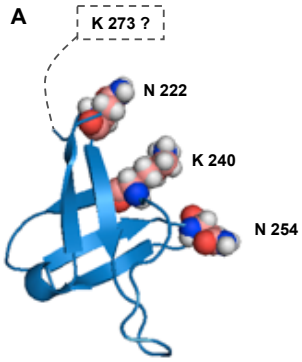


Figure 8: Panel A. Percentage of 2 LTR circles compared to total viral DNA, expressed as a percentage of the catalytic mutant IN A (mutation $D_{116}A$), called IN(-), fixed at 100%. The table under the graphic contains the symbols that reflected the p value of a Student test compared to the parental IN A (top line) or the IN $D_{116}A$ (bottom line). **Panel B.** Percentage of perfect-junction 2LTRc expressed as a function of the total amount of 2LTRc, reported to IN(-), fixed at 100%. The table under the graphic contains the symbols that reflected the p value of a Student test compared to the parental IN A (top line) or the IN $D_{116}A$ (bottom line). The experiments have been repeated at least three times. Error bars are standard déviations.
NS p value>5%, * p value<5%; ** p value<1%; *** p value<0,1%.



B

	Pr55Gag processing *(% of wt A)*	Reverse transcription *(% of wt A)*	Integration normalized by viral DNA *(% of wt A)*
DKDK	72% ± 25%	73% ± 12%	3% ± 2%
	NS	NS	***

Figure 9: Panel A. Representation of the crystallographic structure of the CTD (PDB code: 1IHV). The residues 222, 240, 254 are highlight (273 is part of an unresolved tail). **Panel B.** Pr55Gag processing, reverse transcription and integration normalized by viral DNA for the mutant DKDK, compared to the wt A. Experience has been repeated at least 3 times, SD, Student test at one sample compared to the value of A.
(NS p value>5%, * p value<5%; ** p value<1%; *** p value<0,1%).

Expectedly, the integration competent mutant KQNK, displayed levels of 2LTRc comparable to the parental proteins.

The analysis of the nature of the LTR-LTR junction in 2LTRc reveals a high proportion of unprocessed 3' extremities for the LKLK mutant suggesting a major processing defect, while for the other mutant of the polar residues (TKTK) the proportion of unprocessed 3' ends was comparable to that of the parental IN suggesting that the defects mostly occur at other steps. For the two integration-deficient mutants NQNK and NQNQ, the proportion of unprocessed 3' ends was comparable to that of the $D_{116}A$ mutant, indicating that, besides a nuclear import problem (Figure 8A), these mutants also have an inefficient 3' processing. The integration competent KQNK mutant, as expected, also in this case gave a result comparable to that of the parental proteins (Figure 8B).

The position that characterise the motif was then highlighted on the available structure of the CTD group M (PDB code: 1IHV) showing that the three residues visible on the structure (position 273 is part of an unresolved tail) are aligned (Figure 9A). We supposed that 273 might be aligned with the three other one, besides the N_{222} as residue 270 point toward the N_{222}. We generated the mutant DKDK in order to test if the presence of acidic residues disrupts integration activity. Expectedly, integration is markedly affected with this mutant (3%), rather than Pr55Gag processing and reverse transcription levels are similar to wt A (Figure 9B).

Discussion

For this part, please refer to the thesis, where a part in French discusses the results presented above.

References

Bukovsky, A. & Göttlinger, H. Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease. *J. Virol.* **70,** 6820–6825 (1996).

Buonaguro, L., Tornesello, M. L. & Buonaguro, F. M. Human Immunodeficiency Virus Type 1 Subtype Distribution in the Worldwide Epidemic: Pathogenetic and Therapeutic Implications. *J Virol* **81,** 10209–10219 (2007).

Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P. & Craigie, R. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 3428–3432 (1993).

Busschots, K. *et al.* The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J. Biol. Chem.* **280,** 17841–17847 (2005).

Busschots, K. *et al.* Identification of the LEDGF/p75 binding site in HIV-1 integrase. *J. Mol. Biol.* **365,** 1480–1492 (2007).

Cannon, P. M., Byles, E. D., Kingsman, S. M. & Kingsman, A. J. Conserved sequences in the carboxyl terminus of integrase that are essential for human immunodeficiency virus type 1 replication. *J. Virol.* **70,** 651–657 (1996).

Chen, J. C. *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. U.S.A.* **97,** 8233–8238 (2000).

Chen, A., Weber, I. T., Harrison, R. W. & Leis, J. Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat Ends. *J. Biol. Chem.* **281,** 4173–4182 (2006).

Ciuffi, A. *et al.* A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* **11,** 1287–1289 (2005).

Craigie, R. & Bushman, F. D. HIV DNA Integration. *Cold Spring Harb Perspect Med* **2,** (2012).

Delelis, O., Carayon, K., Saïb, A., Deprez, E. & Mouscadet, J.-F. Integrase and integration: biochemical activities of HIV-1 integrase. *Retrovirology* **5,** 114 (2008).

Devroe, E., Engelman, A. & Silver, P. A. Intracellular transport of human immunodeficiency virus type 1 integrase. *J. Cell. Sci.* **116,** 4401–4408 (2003).

Dobard, C. W., Briones, M. S. & Chow, S. A. Molecular mechanisms by which human immunodeficiency virus type 1 integrase stimulates the early steps of reverse transcription. *J. Virol.* **81,** 10037–10046 (2007).

Dyda, F. *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266,** 1981–1986 (1994).

Gao, K., Buteler, SL., Bushman, F., Human immunodeficiency virus type I integrase: arrangement of protein domains in active cDNA complexes. *EMBO J* **20**: 3565-3576 (2001).

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).

Eijkelenboom, A. P. *et al.* The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Curr. Biol.* **7,** 739–746 (1997).

Engelman, A., Mizuuchi, K. & Craigie, R. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67,** 1211–1221 (1991).

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

Engelman, A., Englund, G., Orenstein, J. M., Martin, M. A. & Craigie, R. Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J Virol* **69,** 2729–2736 (1995).

Esposito, D. & Craigie, R. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* **17,** 5832–5843 (1998).

Galli, A. *et al.* Patterns of Human Immunodeficiency Virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. *J. Virol.* **84,** 7651–7661 (2010).

Gijsbers, R. *et al.* Role of the PWWP domain of lens epithelium-derived growth factor (LEDGF)/p75 cofactor in lentiviral integration targeting. *J. Biol. Chem.* **286,** 41812–41825 (2011).

Heuer, T. S. & Brown, P. O. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry* **36,** 10655–10665 (1997).

Hehl, E. A., Joshi, P., Kalpana, G. V. & Prasad, V. R. Interaction between human immunodeficiency virus type 1 reverse transcriptase and integrase proteins. *J. Virol.* **78,** 5056–5067 (2004).

Hu, W. S. & Temin, H. M. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci. U.S.A.* **87,** 1556–1560 (1990).

Jayappa, K. D., Ao, Z., Yang, M., Wang, J. & Yao, X. Identification of critical motifs within HIV-1 integrase required for importin α3 interaction and viral cDNA nuclear import. *J. Mol. Biol.* **410,** 847–862 (2011).

Johnson, AA., Santos, W., Pais, GC., Marchand, C., Amin, R., Burke,Jr TR., Verdine, G., Pommier, Y. Integration recquires a specific interaction with the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J biol Chem* **281**: 461-467 (2006).

Karki, RG., Tang, Y., Burke, Jr TR., Nicklaus, MC., Model of full-length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design. *J Comput Aided Mol Des* **18**: 739-760 (2004).

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).

Kulkosky, J. & Skalka, A. M. Molecular mechanism of retroviral DNA integration. *Pharmacol. Ther.* **61,** 185–203 (1994).

Larder, B. A., Purifoy, D. J., Powell, K. L. & Darby, G. Site-specific mutagenesis of AIDS virus reverse transcriptase. Nature 327, 716–717 (1987).

Lecossier, D., Bouchonnet, F., Clavel, F. & Hance, A. J. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **300,** 1112 (2003).

Lederle, A. *et al.* Neutralizing Antibodies Inhibit HIV-1 Infection of Plasmacytoid Dendritic Cells by an FcγRIIa Independent Mechanism and Do Not Diminish Cytokines Production. *Scientific Reports* **4,** srep05845 (2014).

Li, X., Krishnan, L., Cherepanov, P. & Engelman, A. Structural biology of retroviral DNA integration. *Virology* **411,** 194–205 (2011).

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).

Lutzke, R. A., Vink, C. & Plasterk, R. H. Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res* **22,** 4125–4131 (1994).

Maillot, B. *et al.* Structural and functional role of INI1 and LEDGF in the HIV-1 preintegration complex. *PLoS ONE* **8,** e60734 (2013).

Michel, F. *et al.* Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. *EMBO J.* **28,** 980–991 (2009).

Passos, D. O. *et al.* Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* **355,** 89–92 (2017).

Pauza, C. D. Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. *Virology* **179,** 886–889 (1990).

Preston, B. D., Poiesz, B. J. & Loeb, L. A. Fidelity of HIV-1 reverse transcriptase. *Science* **242,** 1168–1171 (1988).

Quillent, C., Borman, A. M., Paulous, S., Dauguet, C. & Clavel, F. Extensive regions of pol are required for efficient human immunodeficiency virus polyprotein processing and particle maturation. *Virology* **219,** 29–36 (1996).

Naldini, L. *et al.* In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272,** 263–267 (1996).

Rossolillo, P., Winter, F., Simon-Loriere, E., Gallois-Montbrun, S. & Negroni, M. Retrovolution: HIV-driven evolution of cellular genes and improvement of anticancer drug activation. *PLoS Genet.* **8,** e1002904 (2012).

Sarzotti-Kelsoe M, Bailer RT, Turk E, et al., Optimization and validation of the TZM-bl assay for standardized assessment of neutralizing antibodies against HIV-1. J Immunol Methods**. 409**:131–146 (2014).

Simon-Loriere, E. *et al.* Molecular Mechanisms of Recombination Restriction in the Envelope Gene of the Human Immunodeficiency Virus. *PLOS Pathogens* **5,** e1000418 (2009).

Singh, P. K. *et al.* LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. *Genes Dev.* **29,** 2287–2297 (2015).

Vozzolo, L. *et al.* Gyrase B inhibitor impairs HIV-1 replication by targeting Hsp90 and the capsid protein. *J. Biol. Chem.* **285,** 39314–39328 (2010).

Wang, J.-Y., Ling, H., Yang, W. & Craigie, R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J* **20,** 7333–7343 (2001).

Wilkinson, T. A. *et al.* Identifying and characterizing a functional HIV-1 reverse transcriptase-binding site on integrase. *J. Biol. Chem.* **284,** 7931–7939 (2009).

Woo, J., Robertson, D. L. & Lovell, S. C. Constraints from protein structure and intra-molecular coevolution influence the fitness of HIV-1 recombinants. *Virology* **454,** 34–39 (2014).

Wu, X. *et al.* Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions with the nucleoprotein reverse transcription complex. *J. Virol.* **73,** 2126–2135 (1999).

Zaitseva, L. *et al.* HIV-1 exploits importin 7 to maximize nuclear import of its DNA genome. *Retrovirology* **6,** 11 (2009).

Zheng, R., Jenkins, T. M. & Craigie, R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 13659–13664 (1996).

Zhu, K., Dobard, C. & Chow, S. A. Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase. *J. Virol.* **78,** 5045–5055 (2004).

Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat. Biotechnol.* **15,** 871–875 (1997).

Discussion

Nous avons mis en évidence plusieurs points relevants dans notre présente étude :
1) Un motif conservé composé de deux lysines et deux asparagines ($N_{222}K_{240}N_{254}K_{273}$) au sein du CTD des intégrases du groupe M est essentiel pour l'infectivité.
2) Le motif jouerait un rôle modéré dans l'import nucléaire et un autre, plus important, dans le clivage en 3' des LTRs, ainsi qu'un rôle, non défini, *in vivo*.
3) Certaines positions de ce motif, que ce soit celles où sont présentent les lysines ou les asparagines, sont interchangeables et peuvent être substituées par un acide aminé présentant des propriétés biochimiques similaires (conservation de la polarité et de la charge, N→Q, K→R), sans perturber l'activité de l'IN.
4) Les résidus du motif sont alignés sur un même plan et pourraient former une surface d'interaction (soit avec l'ADN cible, soit inter domaines, soit les deux).

Le domaine C-terminal de l'IN (CTD) est principalement impliqué dans la liaison à l'ADN et à la transcriptase inverse (RT). Plusieurs résidus conservés ont été montrés être importants pour la liaison à l'ADN[58] ($R_{228}R_{231}E_{246}A_{248}R_{263}K_{266}$) et pour l'interaction avec la RT[273,274] ($R_{231}W_{243}G_{247}A_{248}V_{250}I_{251}K_{258}$), mais aussi pour l'import nucléaire[280].
Nos résultats montrent qu'un ensemble de quatre acides aminés ($N_{222}K_{240}N_{254}K_{273}$) non conservés entre les groupes M et O (à l'exception de K_{273}, conservé chez le groupe O mais pas chez l'isolat primaire utilisé, dont la présence d'une glutamine à cette position est atypique) est nécessaire pour l'activité de l'intégrase. En effet, le remplacement des lysines (mutant NQNQ, *Kanja et al.* Figure 4B) ou des asparagines (mutants LKLK, TKTK, *Kanja et al.* Figure 7D) résulte en une perte drastique de l'efficacité d'intégration.

La caractérisation fonctionnelle du motif NKNK suggère qu'il jouerait un rôle dans l'import nucléaire, la transcription inverse et l'intégration, notamment dans la première étape consistant dans le clivage en 3' des LTRs. D'ailleurs, en accord avec nos résultats, l'implication de certains résidus du motif dans des rôles similaires a indépendamment été mise en évidence. Le résidu K_{240} de l'IN a été mis en évidence, en association avec le résidu K_{244}, pour un premier rôle, modéré, dans l'import nucléaire ainsi qu'un autre, plus important, dans la transcription inverse (double mutant K240E-K244E, 21% de la quantité d'ADN viral produite par le sauvage)[325]. Nos résultats montrent que la mutation de la lysine 240 en glutamine affecte l'efficacité de transcription inverse (K240Q, 48% du parent, *Kanja et al.* Figure 3C) et que les deux mutants du motif comportant la lysine 240 mutée (NQNQ et NQNK, *Kanja et al.* Figure 8A), présentent une faible quantité de cercles à 2 LTR, signe d'un

défaut d'import nucléaire. Comme la baisse d'efficacité de transcription inverse de K240Q est bien moins importante que celle observée pour le double mutant K240E-K244E, on peut supposer que dans ce cas, c'est la mutation de la lysine 244, ou la présence d'une charge négative à cette position qui perturbe fortement cette activité. La mutation N222K de l'IN A génère, dans notre système, une IN plus active que le sauvage (N222K, 133% du parent, *Kanja et al.* Figure 3E), bien qu'elle induise une transcription inverse plus faible (N222K, 61% du parent, *Kanja et al.* Figure 3C). Il est possible que ces deux activités se compensent, rendant le virus mutant infectieux, d'ailleurs les virus réplicatifs présentant la substitution N222K dans l'IN ont une infectivité parentale (mutants KQKQ et KQNK, *Kanja et al.* Figure 6), comme déjà décrit dans la littérature[320]. Ces deux résidus (K_{240} et N_{222}) ont été mis en évidence, par spectroscopie à résonnance magnétique nucléaire[274], pour leur implication dans l'interaction IN-RT. Cette implication est secondaire et elle n'a pas été contrôlée par mutagénèse dirigée (comme cela a été fait, en revanche, pour les résidus principaux composant la surface d'interaction, Introduction, partie **II.3.a**). Ainsi, au vu de la baisse modérée de transcription inverse des virus mutants on peut supposer que ces deux résidus ne jouent effectivement pas un rôle principal dans le maintien de l'interaction IN-RT, confirmant les résultats présents dans la littérature, mais qu'ils sont assez proches de la surface d'interaction pour la gêner lorsqu'ils sont mutés. Ainsi, ces deux résidus (K_{240} et N_{222}) affectent peu la transcription inverse lorsqu'ils sont mutés mais jouent un rôle essentiel au sein du motif NKNK pour l'intégration.

Une autre étude a soulevé l'hypothèse, par l'analyse de délétions séquentielles en C-ter de l'IN, que la lysine 273, en association avec les résidus 271 et 272, pourrait affecter la transcription inverse[326], car leur délétion résulte en un taux de transcription inverse de 25% du sauvage. Une autre étude, menée également par l'analyse de délétions séquentielles en C-ter de l'IN, a montré que la délétion jusqu'à la lysine 273 aboutissait à un phénotype de classe I (Introduction, partie **II.3**), c'est à dire qu'elle affecte spécifiquement l'intégration[318]. Nos résultats montrent que la substitution de la lysine 273 par une glutamine affecte de façon modérée la transcription inverse (K273Q, 61% du parent, *Kanja et al.* Figure 3C) mais de façon plus importante l'intégration (K273Q, 21% du parent, *Kanja et al.* Figure 3E). La perte d'efficacité de transcription inverse que nous observons avec la substitution K273Q est moindre comparée à celle de la délétion des résidus de l'IN 271, 272 et 273 , suggérant donc que le résidu 273 ne soit pas impliqué dans le maintien de la transcription inverse. Ainsi, la baisse modérée d'efficacité de transcription inverse pourrait être due à la présence de la glutamine (avec la chaine latérale moins longue et plus encombrée comparée à celle de la lysine) au contact des résidus Y_{271} et G_{272}, potentiellement importants pour assurer le fonctionnement de la transcriptase inverse. On peut donc supposer que le motif identifié

n'est pas directement impliqué dans la transcription inverse mais que les substitutions effectuées pourraient gêner les résidus du CTD relevant pour le bon déroulement de cette étape. Le motif identifié aurait donc principalement un impact sur l'import nucléaire et la première étape de l'intégration. Il faudrait cependant confirmer les défauts observés par d'autres expériences (IN suivi par fluorescence pour l'import nucléaire et test d'intégration concertée *in vitro* pour l'intégration). La substitution K273Q serait donc plutôt impliquée dans un défaut d'intégration que dans la transcription inverse. La lysine 273 a également été identifiée, en association avec l'arginine 269, pour son rôle dans la liaison à l'ARN lors de la morphogénèse de la particule, la mutation en alanine de ces deux résidus basiques résultant en la mauvaise localisation de la RNP (voir dans l'introduction, partie **II.3.b**) au sein des virions[285]. Une mauvaise morphogénèse de la particule, c'est à dire une absence de RNP dans la capside virale, résulterait en un défaut de transcription inverse. Cependant, l'activité parentale en cycle unique et réplicatif de l'un des mutants fonctionnels du motif, avec la lysine 273 mutée en glutamine (KQKQ, *Kanja et al.* Figure 6), suggère que la combinaison de ces mutations ne perturbe pas la liaison à l'ARN, mettant une nouvelle fois en évidence la flexibilité de séquence dont peut faire preuve l'IN. Néanmoins, il faudrait confirmer cette hypothèse en testant plus directement la liaison à l'ARN et la morphogénèse avec d'autres mutants incluant la mutation ponctuelle K273Q. Ainsi, le résidu 273 serait, dans notre étude, relativement peu important pour la transcription inverse et la morphogénèse de la particule mais jouerait un rôle majeur, au sein du motif NKNK, pour l'intégration.

Nous avons ainsi montré que dans un motif comportant deux lysines, les positions auxquelles celles-ci se trouvaient n'avaient que peu d'importance, à l'exception de la combinaison présentant deux lysines en C-ter du motif (NQKK, 20% d'intégration, *Kanja et al.* Figure 4B), démontrant une certaine flexibilité de séquence. D'ailleurs, nous avons également observé que le remplacement des acides aminés du motif par des résidus présentant des propriétés biochimiques similaires (conservation de la polarité et de la charge, N→Q, K→R) n'avait pas d'impact sur l'activité d'intégration de la protéine, bien que certains de ces mutants puissent être qualifiées de classe II puisqu'ils génèrent une baisse d'efficacité dans le clivage des précurseurs polyprotéiques et dans la transcription inverse (NRNR, QKQK, QRQR, *Kanja et al.* Figure 7). Les résidus du motif présents dans la structure cristalline du $CTD_{220-270}$ HIV-1 (PDB code : 1IHV ; acides aminés N_{222}, K_{240} et N_{254}) sont alignés sur un même plan (*Kanja et al.* Figure 9A). Bien que le résidu 273 fasse partie de la queue du CTD absente dans la structure, nous avons supposé qu'il pouvait être aligné avec les trois autres (N_{222}, K_{240} et N_{254}) à côté du N_{222}, car le résidu 270 est à proximité de cette position dans la structure. La possibilité de permutation des acides aminés dans le motif, sans perturber la fonctionnalité, ainsi que l'alignement des résidus suggère que le

motif NKNK forme une surface qui aurait une certaine flexibilité, limitée toutefois par des contraintes structurelles, puisque toutes les positions ne sont pas équivalentes. Lorsqu'une seule lysine et trois acides aminés polaires sont présents dans le motif, deux positions génèrent une baisse d'intégration considérable bien qu'elles conservent une activité basale (222 et 240) (KQNQ, NKNQ, *Kanja et al.* Figure 4B) alors que les positions 254 et 273 (NQKQ, NQNK, *Kanja et al.* Figure 4B) génèrent une integration nulle, similaire à l'intégrase ne présentant aucune lysine dans le motif (NQNQ, *Kanja et al.* Figure 4B). On retrouve d'ailleurs le reflet de ce dernier phénotype avec le motif comportant deux lysines en position 254 et 273 : cette combinaison est la moins fonctionnelle (NQKK, 20% d'intégration, *Kanja et al.* Figure 4B). Ainsi, la présence d'une lysine en position 222 du motif NKNK semblerait pouvoir compenser la mutation en glutamine de K_{240} (KQKQ, KQNK, *Kanja et al.* Figure 4B) et de K_{273} (KKNQ, *Kanja et al.* Figure 4B), puisque ces intégrases (KQKQ, KQNK et KKNQ, Kanja et al. Figure 4B) ont une activité parentale. D'ailleurs, la présence d'au moins deux lysines n'est pas la seule condition pour l'activité puisque lorsque les acides aminés polaires portant un groupement amine dans la chaine latérale (N ou Q) sont remplacés par des acides aminés non polaires (LKLK, *Kanja et al.* Figure 7D) ou polaires mais sans groupement amine dans la chaine latérale (TKTK, *Kanja et al.* Figure 7D), les intégrases mutantes sont défectueuses. La surface formée par ces quatre résidus semblent donc devoir contenir des résidus chargés, portant des groupements aminées dans leur chaine latérale, afin d'assurer l'intégration, laissant supposer que cela pourrait être une surface d'interaction (avec les phosphates de l'ADN ou inter domaines).

La lysine 240 substituée par une alanine, en association avec K236A, a montré générer des virus mutants non infectieux[327]. Comme la lysine 240 a été décrite enclavée dans la complexe héxadécamérique de transfert de brin d'ordre supérieur (voir Introduction partie **II.2.c**) à proximité de E_{35}, alors que ces deux résidus sont exposés dans la structure du tétramère[58], on suppose que la surface formée par le motif NKNK pourrait être impliquée dans une interaction avec le NTD, dont E_{35} fait partie. En effet, la seule présence au sein du motif du groupe M (et de celui du groupe O) d'acides aminés avec un groupement amine dans la chaîne latérale, chargés positivement (K) ou pouvant l'être selon le contexte (N, Q), suggère que les quatre résidus du motif pourraient former, ensemble, une surface d'interaction avec des résidus acides du NTD. Cette surface serait flexible puisque la permutation des positions est possible dans certains cas, sans perturber l'activité. D'ailleurs cette notion se retrouve avec l'analyse des alignements de séquences d'IN du VIH-2, puisqu'un acide aspartique conservé est présent en position 240 alors qu'une glutamine se retrouve en position 35. L'interaction Q-E est donc conservée bien que déplacée.
Comme la présence de lysines, pouvant être remplacées par des arginines, est nécessaire

pour la fonctionnalité du motif, il se pourrait que celui-ci soit également impliqué dans des interactions inter charges avec les phosphates du squelette de la double hélice d'ADN. La mise en évidence des résidus du motif sur la structure du complexe de transfert de brin d'ordre supérieur[58] (PDB non publié, généreusement fourni par D. Lyumkis, Laboraoty of Genetics and Hemsley Center for Genomic Medicine, The Salk Institute for Biological Studies, La Jolla, USA) révèle qu'ils sont positionnés en direction de l'ADN cible mais trop loin pour une interaction. Cependant, la superposition des modèles[231] de tétramères d'IN en liaison avec LEDGF, l'ADN viral et cellulaire des étapes de clivage de l'ADN viral et de transfert de brin indique que les résidus visibles du motif (N_{222}, Q_{240} et N_{254}) sont alignés à proximité de l'ADN cible (Figure **42**). Il se pourrait que le motif NKNK soit impliqué dans la liaison à l'ADN cible, lors d'une transition conformationnelle du complexe d'IN, aboutissant au transfert du brin d'ADN viral à l'ADN cellulaire.

Ce rôle charnière du CTD dans le complexe d'intégration a été mis en évidence pour l'IN d'un autre rétrovirus, MMTV, dont l'intasome présente un arrangement octamérique[236]. L'idée est que les CTD appartenant aux dimères d'IN flanquant le cœur central tétramérique (voir introduction, partie **II.1.d**) permettent d'assurer l'intégration en *trans*. En effet, cette hypothèse s'est forgée sur la base de l'analyse par mutation de l'arginine 240 de l'IN du MMTV (correspondant au résidu E_{246} de l'IN du VIH-1, qui contacte l'extrémité de l'ADN viral), qui assurerait à la fois des contacts inter protomères, et des interactions avec l'ADN viral, essentiels à l'intégration.

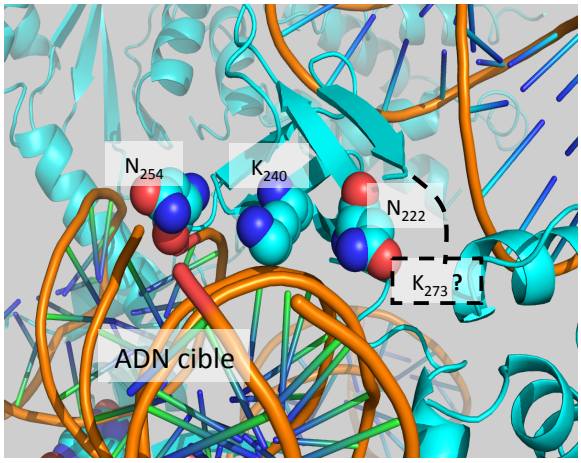


Figure 42 : Superposition des modèles obtenus par cryoéléctromicroscopie du tétramère d'IN en complexe avec LEDGF et l'ADN viral et cellulaire des étapes de clivage des LTR et de transfert de brin. Les acides aminés visibles du motif ont été mis en évidence (N_{222}, K_{240}, N_{254}), la position présumée du K_{273} est indiquée en pointillé. Adapté de[231].

Ainsi, nous avons pu mettre en évidence, pour la première fois, un motif de quatre résidus formant probablement une surface d'interaction (soit avec l'ADN cible, soit inter domaines, soit les deux), relativement flexible et essentielle à l'intégration. Cependant, malgré cette relative flexibilité, la conservation totale du motif au sein du groupe M suggère une importance pour l'infectivité des virus non détectée dans notre système de réplication en

cycle unique. Néanmoins, le test de certains de ces mutants fonctionnels (mutants N222K-K240Q → KQNK et quadruple mutants N222K-K240Q-N254K-K273Q → KQKQ) en cycle réplicatif a démontré que les virus présentaient des taux d'infectivité similaires au sauvage. Le motif mis en évidence aurait donc un rôle important *in vivo*, non mis en évidence dans des systèmes de réplication sur cellules en culture. D'ailleurs, comme une étude sur la fragilité génétique de la capside du VIH-1[328] a montré que des virus avec une fonctionnalité inférieure à 40% in vitro, étaient rarement retrouvés dans la pandémie (<3%), le phénotype de classe II de certains mutants pourrait expliquer la conservation du motif, puisque qu'une diminution des rôles non catalytiques de l'IN résulterait en une baisse de l'infectivité, ne permettant pas aux virus d'être sélectionnés dans la pandémie face aux parentaux, plus fonctionnels.

Conclusions et perspectives

La diversité génétique importante du VIH peut interférer avec la fonctionnalité de ses protéines, puisque l'apparition de mutations peut briser les interactions inter- et intraprotéiques, nécessaires au maintien de l'infectivité. Ces mutations délétères peuvent être contournées par la sélection de mutations compensatoires, reliant ainsi les résidus par un réseau de coévolution. On parle de coévolution des résidus d'une protéine lorsque ces acides aminés évoluent en parallèle, car chacun exerce une pression sélective sur l'autre, affectant ainsi son évolution, dans le but de conserver leur interaction. Ces résidus forment un réseau dynamique d'interactions nécessaire pour le maintien de l'activité de la protéine, et ne sont pas conservés au sein de souches phylogénétiquement distantes, qui ont évolué indépendamment, car les réseaux n'ont pas forcément suivi les mêmes évènements évolutifs, ni subit les mêmes pressions de sélection.
La recombinaison génétique (recombinaison lors de la transcription inverse de deux ARN génomiques viraux différents) peut perturber de tels réseaux en introduisant plusieurs mutations à la fois, rendant la compensation par la sélection de plusieurs mutations successives moins probable. Par ce travail, nous avons caractérisé les contraintes coévolutives qui limitent la diversification de séquence de l'intégrase afin de comprendre les relations structure-fonction entre ses domaines, dans les différentes phases du cycle réplicatif. A cette fin, nous avons construit des chimères entre des intégrases phylogénétiquement distantes, comme celles appartenant aux groupes M et O du VIH-1, dans le but de perturber les réseaux de coévolution. Si ces réseaux sont spécifiques de chaque groupe, la construction des chimères reproduira le résultat obtenu naturellement par recombinaison génétique. L'analyse de la fonctionnalité des virus arborant ces intégrases chimères nous a permis de cartographier plusieurs régions de l'IN impliquées dans le maintien de ses fonctions catalytiques (intégration) ou non catalytiques (transcription inverse, maturation, import nucléaire).

Concernant le rôle de l'intégrase dans la maturation, celui-ci a déjà été mis en évidence par des délétions de la séquence de l'intégrase, comme décrit précédemment, probablement à cause d'une mauvaise dimérisation des Pr160Gag-Pol, qui inhibe l'activation de la protéase. Nous avons montré que certaines chimères A/O étaient défectueuses dans le clivage du précurseur Pr55Gag suggérant l'absence d'autoactivation de la protéase, probablement due à une mauvaise dimérisation des précurseurs Gag-Pol arborant l'intégrase chimère. Nous supposons que la plupart de ces défauts sont provoqués par une structuration différente du domaine IN d'origine O, non compatible avec le reste de Gag-Pol, d'origine M, soulevant l'hypothèse d'une coévolution entre les domaines IN et PR des précurseurs au sein de chaque groupe. D'ailleurs, il semblerait que la perturbation de ce réseau de coévolution (PR-IN) puisse être compensé par le domaine RT puisque lorsque les domaines RT et IN de

Gag-Pol sont tous deux d'origine O au sein du Gag-Pol M, la maturation est parentale, probablement par la compensation du repliement spécifique de l'IN O par celui de la RT O, favorisant la dimérisation et l'autoclivage de la protéase M. Ainsi, nous avons pu mettre en évidence la présence d'un potentiel réseau de coévolution au sein des domaines Pol (PR, RT et IN) du précurseur Gag-Pol, qui ne serait pas conservé entre les groupes M et O.

L'analyse fine de la maturation nous a permis de montrer que le domaine cœur catalytique de l'IN semble être le plus important pour la dimérisation des Pr160Gag-Pol, puisque lorsqu'il est entièrement ou en partie substitué par des régions d'origine O, la maturation est affectée. Le CCD porte la surface de dimérisation[223] entre monomères d'intégrase mature, mais il se pourrait qu'elle soit aussi impliquée dans la dimérisation du précurseur Gag-Pol. Il est possible qu'un repliement différent dû à des portions de séquences d'origines différentes dans ce domaine ou dans une autre région de la protéine puisse perturber cette surface ou générer une surface alternative, non compatible avec le reste du précurseur. L'absence de dimérisation des précurseurs Gag-Pol peut affecter l'autoactivation de la protéase et ainsi inhiber la maturation.

Mon travail a permis de mettre en évidence que certains résidus isolés (Figure **41**) pourraient avoir un rôle structural dans le maintien de la surface de dimérisation. Cependant, les défauts observés sont peu marqués et il n'est pas possible, dans notre système en cycle unique, de définir s'ils sont de simples variations négligeables pour l'infectivité des virus ou s'ils peuvent perturber suffisamment l'infectivité, de telle sorte que ces polymorphismes ne seront pas sélectionnés dans la pandémie car moins fonctionnels que les parentaux. En effet, la plupart de ces virus, avec un taux de clivage de Pr55Gag d'environ 60%, présentent des efficacités de transcription inverse parentale dans notre système d'étude (72/106, Q137H, S153A, S195T).

Par la suite, il serait très intéressant de tester nos chimères et mutants ponctuels en cycle réplicatifs, afin de déterminer si une efficacité de clivage des précurseurs polyprotéiques de 60% est suffisante ou non pour assurer l'infectivité des virus. Comme nous supposons que la dimérisation du domaine CCD pourrait être importante pour la maturation, il serait également très intéressant de déterminer les structures cristallographiques des domaines CCD[206] chimères, mutés et d'origine O, afin d'évaluer s'ils présentent la même structure que le CCD d'origine M, s'ils sont présents sous forme dimérique dans le cristal, et si ce n'est pas le cas, de déterminer qu'elles interactions perturberaient la structure de la surface de dimérisation.

L'interaction entre l'IN et la RT[273,274] est nécessaire au bon déroulement de la transcription inverse et assure la stabilisation et le placement correct de l'amorce $tRNA^{Lys,3}$ sur la matrice[275]. Nous avons montré dans notre étude que de nombreuses chimères au sein des trois domaines de l'IN et de nombreux résidus présents dans ces régions affectaient la

fonctionnalité de la transcriptase inverse de façon plus ou moins importante lorsqu'ils sont mutés (13 résidus, Figure **41**). La plupart de ces chimères/résidus affectent une étape précoce de la transcription inverse, suggérant qu'ils pourraient perturber l'interaction IN-RT et ainsi, inhiber le rôle de l'IN sur l'initiation de la transcription inverse. Les mutations dans le CTD doivent probablement affecter directement la surface d'interaction avec la RT ou être eux mêmes impliqués dans l'interaction. Par contre, les résidus présents dans les domaines NTD et CCD doivent affecter indirectement le maintien de l'interaction au niveau du CTD, probablement en déstabilisant le repliement général de la protéine, qui rendrait moins accessible ou inaccessible le domaine CTD. Ainsi, nous avons pu mettre en évidence plusieurs régions/résidus de l'IN qui sembleraient jouer un rôle dans l'activité de la transcriptase inverse par le maintien du repliement fonctionnel de l'IN, favorable pour son interaction avec la RT. L'IN présenterait donc des réseaux de coévolution non conservés entre les groupes M et O qui seraient importants pour l'activité de la RT, soulevant ainsi l'idée d'une éventuelle coévolution entre la RT et l'IN. Par la suite, on pourrait envisager de purifier les domaines CCD-CTD (mutés et chimères) et analyser leur capacité d'interaction avec la RT par résonnance plasmonique[274] ainsi que leur structures crystallographiques[225] pour les comparer à celle des parents et évaluer si la surface d'interaction du CTD est modifiée/déplacée. Plusieurs résidus (Figure **41**) ont montré présenter à la fois un défaut dans la transcription inverse et dans l'intégration lorsqu'ils sont substitués par la séquence d'origine O. On ne peut donc pas exclure que les défauts observés avec ces substitutions puissent être dus à un problème survenant avant le processus de transcription inverse, comme un défaut dans la décapsidation ou dans la morphogénèse de la particule puisque, comme décrit précédemment, l'IN joue également un rôle dans ces étapes. Il faudrait donc contrôler, par la suite, la liaison à l'ARN et la bonne morphogénèse des virus[285] portants les IN mutantes et chimères, ainsi que le bon déroulement de la décapsidation[329].

La construction de chimères ainsi que la caractérisation des résidus impliqués dans le défaut de fonctionnalité nous a permis de mettre en évidence plusieurs acides aminés, tout le long de l'intégrase dont la mutation a un impact sévère sur l'intégration en elle-même (Figure **41**). La mise en évidence de ces résidus sur les structures disponibles des domaines et la concordance avec les données de la littérature, nous a amené à supposer que les acides aminés identifiés avaient majoritairement un rôle dans les interactions structurelles permettant la multimérisation de l'IN et leur substitution provoquerait des défauts multiples selon la région de l'intégrase affectée (défaut de liaison au cofacteur LEDGF, défaut de liaison aux facteurs cellulaires permettant l'import nucléaire, défaut de liaison à l'ADN cible). Par la suite, il faudrait confirmer les défauts observés d'import nucléaire et purifier les intégrases d'intérêt pour tester, *in vitro,* leur capacité à lier les facteurs cellulaires permettant

l'import nucléaire, ainsi que son cofacteur LEDGF[323]. Il serait également intéressant de contrôler *in vitro*, sur ces même intégrases purifiées, leur capacité de clivage de l'ADN viral et de transfert de brin par un test d'intégration concertée[203]. Enfin, il faudrait évaluer la capacité des IN d'intérêt à multimériser par des analyses de chromatographie à exclusion de taille[323]. De plus, comme nous supposons que la plupart des résidus identifiés pourraient avoir un impact sur le repliement de l'IN lorsqu'ils sont mutés, il faudrait déterminer la structure cristallographique des NTD-CCD et CCD-CTD[225] des intégrases mutées et d'origine O, ce qui nous permettrait peut être d'identifier les contraintes structurelles qui perturbent le repliement de l'IN A.

Ce travail nous a également permis de mettre en évidence un motif de quatre résidus au sein du CTD du groupe M ($N_{222}K_{240}N_{254}K_{273}$) essentiel pour l'intégration. Nous avons notamment pu observer que les lysines en positions 240 et 273 de ce motif étaient nécessaires à l'activité (Figure **41**), tout comme les asparagines en positions 222 et 254. Nos résultats laissent à penser que ce motif pourrait constituer une surface d'interaction (soit avec l'ADN cible, soit inter domaines, soit les deux), relativement flexible, puisque plusieurs permutations des positions des résidus sont possibles sans perturber l'activité. La fonctionnalité des motifs mutés en cycle multiples nous a d'ailleurs amené à supposer que ce motif aurait un rôle important *in vivo*, non détecté dans des systèmes d'infection sur cellules, qui expliquerait sa conservation au sein du groupe M. En effet, le contexte des cellules utilisées (HEK293T, CEM-SS, TZM-bL) n'est peut être pas suffisant pour comprendre le rôle du motif, il serait donc très intéressant de tester nos virus en infection sur d'autres cellules (macrophages ou monocytes par exemple). La découverte d'une telle surface essentielle à l'activité et flexible au sein du CTD pourrait être le signe que l'IN maintient sa structure et sa fonctionnalité par des régions charnières et flexibles qui établissent des interactions coévolutives, non conservées entre les groupes M et O.

Nous avons montré par notre approche expérimentale, que plusieurs résidus de l'IN A sont soumis à des contraintes évolutives (puisque non conservés entre M et O mais essentiels à l'activité) afin de maintenir ses rôles catalytiques et non catalytiques. On pourrait penser que les résidus identifiés participent, comme pour le motif du CTD, à des interactions importantes, pouvant supporter une certaine flexibilité qui permet de maintenir la structure fonctionnelle de l'IN tout en autorisant une certaine variabilité de séquence.

En conclusion, ce travail de thèse a montré que plusieurs régions/résidus de l'IN ont coévolué de façon indépendante entre les groupes M et O. Comme les résidus responsables de la catalyse ou des interactions avec les cofacteurs (viraux et cellulaires) sont

principalement conservés[321], il semblerait que les réseaux de coévolution permettent principalement d'assurer des interactions structurelles favorisant le repliement favorable de l'IN pour son activité catalytique et ses autres rôles dans le cycle infectieux. Ceci suggère que les intégrases de ces deux groupes auraient un repliement différent. Pour rappel, la prévalence des virus de groupe O n'est pas négligeable par rapport à la pandémie, contrairement à celles des virus des groupes N et P (près de 100 000 individus infectés pour le groupe O contre 20 et 2 pour N et P)[172,173]. De plus, comme des recombinants M/O ont été mis en évidence[189-191], il serait très intéressant de plus amplement étudier les réseaux de coévolution non conservés entre les intégrases de groupes M et O. En effet, caractériser ces relations coévolutives pourrait permettre d'identifier de nouvelles interactions, importantes pour la fonctionnalité de la protéine, ainsi que de comprendre le fonctionnement des IN de groupe O, qui semble différentes des IN M.

Bibliographie

1. ONUSIDA - AIDS-by-the-numbers-2016.
2. Curran, J.W., Jaffe, H.W. AIDS: the Early Years and CDC's Response, *MMWR Morb. Mortal. Wkly. Rep,* **60(04)**, 64-69 (2011).
3. Center for Disease Control (CDC). Update on Acquired Immune Deficiency Syndrome (AIDS) among Patients with Hemophilia A. *MMWR Morb. Mortal. Wkly. Rep.* **31**, 507-513 (1982).
4. Popovic, M., Reitz, M.S., Sarngadharan, M.G., Robert-Guroff, M., Kalyanaraman, V.S., Nakao, Y., Miyoshi, I., Minowada, J., Yoshida, M., Ito, Y., et al. The virus of Japanese adult T-cell leukaemia is a member of the human T-cell leukaemia virus group. *Nature* **300**, 63–66 (1982).
5. Gallo, R.C., Blattner, W.A., Reitz, M.S., and Ito, Y. HTLV: the virus of adult T-cell leukaemia in Japan and elsewhere. *Lancet* **1**, 683 (1982).
6. Barre-Sinoussi, F., Chermann, J., Rey, F., Nugeyre, M., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**, 868–871 (1983).
7. Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., and Safai, B. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**, 500–503 (1984).
8. Montagnier, L., Chermann, J.C., Barré-Sinoussi, F., Klatzmann, D., Wain-Hobson, S., Alizon, M., Clavel, F., Brun-Vezinet, F., Vilmer, E., and Rouzioux, C. Lymphadenopathy associated virus and its etiological role in AIDS. *Int. Symp. Princess Takamatsu Cancer Res. Fund* **15**, 319–331 (1984).
9. Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M., Santos-Ferreira, M., Laurent, A., Dauguet, C., Katlama, C., Rouzioux, C., et al. Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**, 343–346 (1986).
10. ICTV Virus Taxonomy 2016. Available at: https://talk.ictvonline.org/taxonomy/
11. Benit, L., Dessen, P., and Heidmann, T. Identification, Phylogeny, and Evolution of Retroviral Elements Based on Their Envelope Genes. *Journal of Virology* **75**, 11709–11719 (2001).
12. Weiss, R.A. The discovery of endogenous retroviruses. *Retrovirology* **3**, 67 (2006).
13. Foley, B., Leitner, T., Apetrei, C., Hahn, B., Mizrachi, I., Mullins, J., Rambaut, A., Wolinsky, S., and Korber, B. *HIV Sequence Compendium 2016*. *Los Alamos National Laboratory*, 1-445 (2016).
14. Darlix, J.-L., Gabus, C., Nugeyre, M.-T., Clavel, F., and Barré-Sinoussi, F. Cis elements and Trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *Journal of Molecular Biology* **216**, 689–699 (1990).
15. Marquet, R., Paillart, J.C., Skripkin, E., Ehresmann, C., and Ehresmann, B. Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucleic Acids Res* **22**, 145–151 (1994).
16. Peterlin, B.M., and Trono, D. Hide, shield and strike back: how HIV-infected cells avoid immune eradication. *Nature Reviews Immunology* **3**, 97–107 (2003).
17. Stein, B.S., and Engleman, E.G. Intracellular processing of the gp160 HIV-1 envelope precursor. Endoproteolytic cleavage occurs in a cis or medial compartment of the Golgi complex. *J. Biol. Chem.* **265**, 2640–2649 (1990).
18. McCune, J.M., Rabin, L.B., Feinberg, M.B., Lieberman, M., Kosek, J.C., Reyes, G.R., and Weissman, I.L. Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell* **53**, 55–67 (1988).
19. Das, A.T., Koken, S.E., Essink, B.B., van Wamel, J.L., and Berkhout, B. Human immunodeficiency virus uses tRNA(Lys,3) as primer for reverse transcription in HeLa-CD4+ cells. *FEBS Lett*. **341**, 49–53 (1994).
20. Resnick, R., Omer, C.A., and Faras, A.J. Involvement of retrovirus reverse transcriptase-associated RNase H in the initiation of strong-stop (+) DNA synthesis and the generation of the long terminal repeat. *J. Virol*. **51**, 813–821 (1984).
21. Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.C., and Montagnier, L. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **312**, 767–768 (1984).
22. Dalgleish, A.G., Beverley, P.C., Clapham, P.R., Crawford, D.H., Greaves, M.F., and Weiss, R.A. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**, 763–767 (1984).
23. Trkola, A., Dragic, T., Arthos, J., Binley, J.M., Olson, W.C., Allaway, G.P., Cheng-Mayer, C., Robinson, J., Maddon, P.J., and Moore, J.P. CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. Nature *384*, 184–187 (1996).
24. Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R.E., Hill, C.M. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661–666 (1996).

25. Zhu, P., Liu, J., Bess, J., Chertova, E., Lifson, J.D., Grisé, H., Ofek, G.A., Taylor, K.A., and Roux, K.H. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441**, 847–852 (2006).
26. Bernstein, H.B., Tucker, S.P., Hunter, E., Schutzbach, J.S., and Compans, R.W. Human immunodeficiency virus type 1 envelope glycoprotein is modified by O-linked oligosaccharides. *J. Virol*. **68**, 463–468 (1994).
27. Starcich, B.R., Hahn, B.H., Shaw, G.M., McNeely, P.D., Modrow, S., Wolf, H., Parks, E.S., Parks, W.P., Josephs, S.F., and Gallo, R.C. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* **45**, 637–648 (1986).
28. Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659 (1998).
29. Cardozo, T., Kimura, T., Philpott, S., Weiser, B., Burger, H., and Zolla-Pazner, S. Structural basis for coreceptor selectivity by the HIV type 1 V3 loop. *AIDS Res. Hum. Retroviruses* **23**, 415–426 (2007).
30. Bartesaghi, A., Merk, A., Borgnia, M.J., Milne, J.L.S., and Subramaniam, S. Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nat. Struct. Mol. Biol*. **20**, 1352–1357 (2013).
31. Tran, E.E.H., Borgnia, M.J., Kuybeda, O., Schauder, D.M., Bartesaghi, A., Frank, G.A., Sapiro, G., Milne, J.L.S., and Subramaniam, S. Structural mechanism of trimeric HIV-1 envelope glycoprotein activation. *PLoS Pathog*. **8**, e1002797 (2012).
32. Chan, D.C., Fass, D., Berger, J.M., and Kim, P.S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263–273 (1997).
33. Shang, L., Yue, L., and Hunter, E. Role of the membrane-spanning domain of human immunodeficiency virus type 1 envelope glycoprotein in cell-cell fusion and virus infection. *J. Virol*. **82**, 5417–5428 (2008).
34. Bukrinskaya, A. HIV-1 matrix protein: a mysterious regulator of the viral life cycle. *Virus Res*. **124**, 1–11 (2007).
35. Hill, C.P., Worthylake, D., Bancroft, D.P., Christensen, A.M., and Sundquist, W.I. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proc. Natl. Acad. Sci. U.S.A*. **93**, 3099–3104 (1996).
36. Dorfman, T., Mammano, F., Haseltine, W.A., and Göttlinger, H.G. Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. *J. Virol.* **68**, 1689–1696 (1994).
37. Spearman, P., Wang, J.J., Vander Heyden, N., and Ratner, L. Identification of human immunodeficiency virus type 1 Gag protein domains essential to membrane binding and particle assembly. *J. Virol.* **68**, 3232–3242 (1994).
38. Tedbury, P.R., Novikova, M., Ablan, S.D., and Freed, E.O. Biochemical evidence of a role for matrix trimerization in HIV-1 envelope glycoprotein incorporation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E182-190 (2016).
39. Yu, X., Yuan, X., Matsuda, Z., Lee, T.H., and Essex, M. The matrix protein of human immunodeficiency virus type 1 is required for incorporation of viral envelope protein into mature virions. *J. Virol*. **66**, 4966–4971 (1992).
40. Borman, S. A. Building HIV's Curvaceous Coat | June 22, 2009 Issue - Vol. 87 Issue 25 | Chemical & Engineering News. Available at: http://cen.acs.org/articles/87/i25/Building-HIVs-Curvaceous-Coat.html. (Accessed: 11th May 2017)
41. Gres, A. T. *et al.* STRUCTURAL VIROLOGY. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **349,** 99–103 (2015).
42. Briggs, J. A. G. *et al.* The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions. *Structure* **14,** 15–20 (2006).
43. Clavel, F. & Orenstein, J. M. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J. Virol.* **64,** 5230–5234 (1990).
44. Aldovini, A. & Young, R. A. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* **64,** 1920–1926 (1990).
45. Stys, D., Blaha, I. & Strop, P. Structural and functional studies in vitro on the p6 protein from the HIV-1 gag open reading frame. *Biochim. Biophys. Acta* **1182,** 157–161 (1993).
46. Garrus, J. E. *et al.* Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107,** 55–65 (2001).

47. Göttlinger, H. G., Dorfman, T., Sodroski, J. G. & Haseltine, W. A. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A* **88,** 3195–3199 (1991).
48. Kohl, N. E. *et al.* Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U.S.A.* **85,** 4686–4690 (1988).
49. Miller, M., Jaskólski, M., Rao, J. K., Leis, J. & Wlodawer, A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* **337,** 576–579 (1989).
50. Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H. & Poljak, R. J. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie* **73,** 1391–1396 (1991).
51. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256,** 1783–1790 (1992).
52. Jacobo-Molina, A. *et al.* Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 6320–6324 (1993).
53. Wang, J. *et al.* Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer. *Proc Natl Acad Sci U S A* **91,** 7242–7246 (1994).
54. Jaeger, J., Restle, T. & Steitz, T. A. The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *EMBO J* **17,** 4535–4542 (1998).
55. Bushman, F. D. & Craigie, R. Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. *Proc. Natl. Acad. Sci. U.S.A.* **88,** 1339–1343 (1991).
56. Bowerman, B., Brown, P. O., Bishop, J. M. & Varmus, H. E. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev.* **3,** 469–478 (1989).
57. Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P. & Craigie, R. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 3428–3432 (1993).
58. Passos, D. O. *et al.* Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* **355,** 89–92 (2017).
59. Dingwall, C. *et al.* Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **86,** 6925–6929 (1989).
60. Dingwall, C. *et al.* HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. *EMBO J.* **9,** 4145–4153 (1990).
61. Cujec, T. P. *et al.* The human immunodeficiency virus transactivator Tat interacts with the RNA polymerase II holoenzyme. *Mol. Cell. Biol.* **17,** 1817–1823 (1997).
62. Taube, R., Fujinaga, K., Wimmer, J., Barboric, M. & Peterlin, B. M. Tat transactivation: a model for the regulation of eukaryotic transcriptional elongation. *Virology* **264,** 245–253 (1999).
63. Meyer, B. E. & Malim, M. H. The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. *Genes Dev.* **8,** 1538–1547 (1994).
64. Kalland, K. H., Szilvay, A. M., Brokstad, K. A., Saetrevik, W. & Haukenes, G. The human immunodeficiency virus type 1 Rev protein shuttles between the cytoplasm and nuclear compartments. *Mol. Cell. Biol.* **14,** 7436–7444 (1994).
65. Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, I. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90,** 1051–1060 (1997).
66. Fisher, A. G. *et al.* The sor gene of HIV-1 is required for efficient virus transmission in vitro. *Science* **237,** 888–893 (1987).
67. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418,** 646–650 (2002).
68. Feng, Y., Baig, T. T., Love, R. P. & Chelico, L. Suppression of APOBEC3-mediated restriction of HIV-1 by Vif. *Front Microbiol* **5,** 450 (2014).
69. Popov, S., Rexach, M., Ratner, L., Blobel, G. & Bukrinsky, M. Viral protein R regulates docking of the HIV-1 preintegration complex to the nuclear pore complex. *J. Biol. Chem.* **273,** 13347–13352 (1998).
70. Mahalingam, S. *et al.* Identification of Residues in the N-Terminal Acidic Domain of HIV-1 Vpr Essential for Virion Incorporation. *Virology* **207,** 297–302 (1995).
71. Le Rouzic, E. & Benichou, S. The Vpr protein from HIV-1: distinct roles along the viral life cycle. *Retrovirology* **2,** 11 (2005).
72. Strebel, K., Klimkait, T., Maldarelli, F. & Martin, M. A. Molecular and biochemical analyses of human immunodeficiency virus type 1 vpu protein. *J. Virol.* **63,** 3784–3791 (1989).

73. Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451,** 425–430 (2008).
74. Bour, S., Geleziunas, R. & Wainberg, M. A. The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. *Microbiol. Rev.* **59,** 63–93 (1995).
75. Garcia, J. V. & Miller, A. D. Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature* **350,** 508–511 (1991).
76. Schwartz, O. *et al.* Human immunodeficiency virus type 1 Nef induces accumulation of CD4 in early endosomes. *J. Virol.* **69,** 528–533 (1995).
77. Kestler, H. W. *et al.* Importance of the nef gene for maintenance of high virus loads and for development of AIDS. *Cell* **65,** 651–662 (1991).
78. Basmaciogullari, S. & Pizzato, M. The activity of Nef on HIV-1 infectivity. *Front Microbiol* **5,** (2014).
79. Saphire, A. C., Bobardt, M. D., Zhang, Z., David, G. & Gallay, P. A. Syndecans serve as attachment receptors for human immunodeficiency virus type 1 on macrophages. *J. Virol.* **75,** 9187–9200 (2001).
80. Cicala, C. *et al.* The integrin alpha4beta7 forms a complex with cell-surface CD4 and defines a T-cell subset that is highly susceptible to infection by HIV-1. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 20877–20882 (2009).
81. Geijtenbeek, T. B. *et al.* DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100,** 587–597 (2000).
82. McDougal, J. S. *et al.* Binding of the human retrovirus HTLV-III/LAV/ARV/HIV to the CD4 (T4) molecule: conformation dependence, epitope mapping, antibody inhibition, and potential for idiotypic mimicry. *J. Immunol.* **137,** 2937–2944 (1986).
83. Berger, E. A. HIV entry and tropism. When one receptor is not enough. *Adv. Exp. Med. Biol.* **452,** 151–157 (1998).
84. Bleul, C. C., Wu, L., Hoxie, J. A., Springer, T. A. & Mackay, C. R. The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. *Proc Natl Acad Sci U S A* **94,** 1925–1930 (1997).
85. Esté, J. A. *et al.* Shift of Clinical Human Immunodeficiency Virus Type 1 Isolates from X4 to R5 and Prevention of Emergence of the Syncytium-Inducing Phenotype by Blockade of CXCR4. *J Virol* **73,** 5577–5585 (1999).
86. Weissenhorn, W. *et al.* Assembly of a rod-shaped chimera of a trimeric GCN4 zipper and the HIV-1 gp41 ectodomain expressed in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 6065–6069 (1997).
87. Wilen, C. B., Tilton, J. C. & Doms, R. W. HIV: cell binding and entry. *Cold Spring Harb Perspect Med* **2,** (2012).
88. 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).
89. Rasaiyaah, J. *et al.* HIV-1 evades innate immune recognition through specific co-factor recruitment. *Nature* **503,** 402–405 (2013).
90. Bukrinsky, M. I. *et al.* Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc Natl Acad Sci U S A* **90,** 6125–6129 (1993).
91. Cosnefroy, O., Murray, P. J. & Bishop, K. N. HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology* **13,** (2016).
92. Colgan, J., Yuan, H. E., Franke, E. K. & Luban, J. Binding of the human immunodeficiency virus type 1 Gag polyprotein to cyclophilin A is mediated by the central region of capsid and requires Gag dimerization. *J Virol* **70,** 4299–4310 (1996).
93. Liu, C. *et al.* Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nature Communications* **7,** ncomms10714 (2016).
94. Ambrose, Z. & Aiken, C. HIV-1 Uncoating: Connection to Nuclear Entry and Regulation by Host Proteins. *Virology* **0,** 371–379 (2014).
95. Mak, J. & Kleiman, L. Primer tRNAs for reverse transcription. *J. Virol.* **71,** 8087–8095 (1997).
96. Kleiman, L., Jones, C. P. & Musier-Forsyth, K. Formation of the tRNALys packaging complex in HIV-1. *FEBS Lett.* **584,** 359–365 (2010).
97. Javanbakht, H. *et al.* The interaction between HIV-1 Gag and human lysyl-tRNA synthetase during viral assembly. *J. Biol. Chem.* **278,** 27644–27651 (2003).
98. Zaitseva, L., Myers, R. & Fassati, A. tRNAs promote nuclear import of HIV-1 intracellular reverse transcription complexes. *PLoS Biol.* **4,** e332 (2006).
99. Telesnitsky, A. & Goff, S. P. in *Retroviruses* (eds. Coffin, J. M., Hughes, S. H. & Varmus, H. E.) (Cold Spring Harbor Laboratory Press, 1997).

100. Levin, J. G., Mitra, M., Mascarenhas, A. & Musier-Forsyth, K. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol* **7,** 754–774 (2010).
101. van Wamel, J. L. & Berkhout, B. The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly. *Virology* **244,** 245–251 (1998).
102. Balasubramaniam, M. & Freed, E. O. New Insights into HIV Assembly and Trafficking. *Physiology (Bethesda)* **26,** 236–251 (2011).
103. Yamashita, M., Perez, O., Hope, T. J. & Emerman, M. Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. *PLoS Pathog.* **3,** 1502–1510 (2007).
104. König, R. *et al.* Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* **135,** 49–60 (2008).
105. Brass, A. L. *et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319,** 921–926 (2008).
106. 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).
107. McDonald, D. *et al.* Visualization of the intracellular behavior of HIV in living cells. *J. Cell Biol.* **159,** 441–452 (2002).
108. Schaller, T. *et al.* HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog.* **7,** e1002439 (2011).
109. Lee, K. *et al.* Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe* **7,** 221–233 (2010).
110. Lusic, M. & Siliciano, R. F. Nuclear landscape of HIV-1 infection and integration. *Nat Rev Micro* **15,** 69–82 (2017).
111. Kao, S. Y., Calman, A. F., Luciw, P. A. & Peterlin, B. M. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* **330,** 489–493 (1987).
112. Pereira, L. A., Bentley, K., Peeters, A., Churchill, M. J. & Deacon, N. J. SURVEY AND SUMMARY A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Res* **28,** 663–668 (2000).
113. Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H. & Jones, K. A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92,** 451–462 (1998).
114. Harrich, D., Hsu, C., Race, E. & Gaynor, R. B. Differential growth kinetics are exhibited by human immunodeficiency virus type 1 TAR mutants. *J. Virol.* **68,** 5899–5910 (1994).
115. Purcell, D. F. & Martin, M. A. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J. Virol.* **67,** 6365–6378 (1993).
116. Schwartz, S., Felber, B. K., Benko, D. M., Fenyö, E. M. & Pavlakis, G. N. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J Virol* **64,** 2519–2529 (1990).
117. Cullen, B. R. Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology* **249,** 203–210 (1998).
118. Farjot, G., Sergeant, A. & Mikaélian, I. A new nucleoporin-like protein interacts with both HIV-1 Rev nuclear export signal and CRM-1. *J. Biol. Chem.* **274,** 17309–17317 (1999).
119. Neville, M., Stutz, F., Lee, L., Davis, L. I. & Rosbash, M. The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Curr. Biol.* **7,** 767–775 (1997).
120. Jacks, T. *et al.* Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **331,** 280–283 (1988).
121. Schubert, U. *et al.* CD4 glycoprotein degradation induced by human immunodeficiency virus type 1 Vpu protein requires the function of proteasomes and the ubiquitin-conjugating pathway. *J. Virol.* **72,** 2280–2288 (1998).
122. Fujita, K., Omura, S. & Silver, J. Rapid degradation of CD4 in cells expressing human immunodeficiency virus type 1 Env and Vpu is blocked by proteasome inhibitors. *J. Gen. Virol.* **78 (Pt 3),** 619–625 (1997).
123. Sundquist, W. I. & Krausslich, H.-G. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harbor Perspectives in Medicine* **2,** a006924–a006924 (2012).
124. Jouvenet, N. *et al.* Plasma membrane is the site of productive HIV-1 particle assembly. *PLoS Biol.* **4,** e435 (2006).
125. Ono, A., Ablan, S. D., Lockett, S. J., Nagashima, K. & Freed, E. O. Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 14889–14894 (2004).

126. Hill, C. P., Worthylake, D., Bancroft, D. P., Christensen, A. M. & Sundquist, W. I. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 3099–3104 (1996).
127. Wright, E. R. *et al.* Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *EMBO J.* **26,** 2218–2226 (2007).
128. Jowett, J. B. M., Hockley, D. J., Nermut, M. V. & Jones, I. M. Distinct signals in human immunodeficiency virus type 1 Pr55 necessary for RNA binding and particle formation. *Journal of General Virology* **73,** 3079–3086 (1992).
129. Yu, X., Yuan, X., McLane, M. F., Lee, T. H. & Essex, M. Mutations in the cytoplasmic domain of human immunodeficiency virus type 1 transmembrane protein impair the incorporation of Env proteins into mature virions. *J. Virol.* **67,** 213–221 (1993).
130. Martin-Serrano, J., Zang, T. & Bieniasz, P. D. HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nat. Med.* **7,** 1313–1319 (2001).
131. Garrus, J. E. *et al.* Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107,** 55–65 (2001).
132. Strack, B., Calistri, A., Craig, S., Popova, E. & Göttlinger, H. G. AIP1/ALIX Is a Binding Partner for HIV-1 p6 and EIAV p9 Functioning in Virus Budding. *Cell* **114,** 689–699 (2003).
133. Hurley, J. H. & Hanson, P. I. Membrane budding and scission by the ESCRT machinery: it's all in the neck. *Nat. Rev. Mol. Cell Biol.* **11,** 556–566 (2010).
134. Viral Zone - HIV home page. Available at: http://viralzone.expasy.org/4976.
135. Cooper, J. B. Aspartic proteinases in disease: a structural perspective. *Curr Drug Targets* **3,** 155–173 (2002).
136. Tang, C., Louis, J. M., Aniana, A., Suh, J.-Y. & Clore, G. M. Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. *Nature* **455,** 693–696 (2008).
137. Kräusslich, H. G., Traenckner, A. M. & Rippmann, F. Expression and characterization of genetically linked homo- and hetero-dimers of HIV proteinase. *Adv. Exp. Med. Biol.* **306,** 417–428 (1991).
138. Prabu-Jeyabalan, M., Nalivaika, E. & Schiffer, C. A. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure* **10,** 369–381 (2002).
139. Pettit, S. C. *et al.* The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.* **68,** 8017–8027 (1994).
140. Rosa, A. *et al.* HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature* **526,** 212–217 (2015).
141. Usami, Y., Wu, Y. & Göttlinger, H. G. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature* **526,** 218–223 (2015).
142. Neil, S. & Bieniasz, P. Human immunodeficiency virus, restriction factors, and interferon. *J. Interferon Cytokine Res.* **29,** 569–580 (2009).
143. Stremlau, M., Song, B., Javanbakht, H., Perron, M. & Sodroski, J. Cyclophilin A: an auxiliary but not necessary cofactor for TRIM5alpha restriction of HIV-1. *Virology* **351,** 112–120 (2006).
144. Laguette, N. *et al.* SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474,** 654–657 (2011).
145. Lahouassa, H. *et al.* SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat. Immunol.* **13,** 223–228 (2012).
146. Lecossier, D., Bouchonnet, F., Clavel, F. & Hance, A. J. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **300,** 1112 (2003).
147. Conticello, S. G., Harris, R. S. & Neuberger, M. S. The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr. Biol.* **13,** 2009–2013 (2003).
148. Yu, X. *et al.* Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302,** 1056–1060 (2003).
149. Klimkait, T., Strebel, K., Hoggan, M. D., Martin, M. A. & Orenstein, J. M. The human immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release. *J. Virol.* **64,** 621–629 (1990).
150. Bour, S. & Strebel, K. The human immunodeficiency virus (HIV) type 2 envelope protein is a functional complement to HIV type 1 Vpu that enhances particle release of heterologous retroviruses. *J. Virol.* **70,** 8285–8300 (1996).
151. Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451,** 425–430 (2008).

152. Ross, T. M., Oran, A. E. & Cullen, B. R. Inhibition of HIV-1 progeny virion release by cell-surface CD4 is relieved by expression of the viral Nef protein. *Curr. Biol.* **9,** 613–621 (1999).
153. Collins, K. L., Chen, B. K., Kalams, S. A., Walker, B. D. & Baltimore, D. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* **391,** 397–401 (1998).
154. Veillette, M. *et al.* Interaction with cellular CD4 exposes HIV-1 envelope epitopes targeted by antibody-dependent cell-mediated cytotoxicity. *J. Virol.* **88,** 2633–2644 (2014).
155. Hatziioannou, T. & Evans, D. T. Animal models for HIV/AIDS research. *Nat Rev Micro* **10,** 852–867 (2012).
156. Fiebig, E. W. *et al.* Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS* **17,** 1871–1879 (2003).
157. Simon, V., Ho, D. D. & Karim, Q. A. HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. *Lancet* **368,** 489–504 (2006).
158. An, P. & Winkler, C. A. Host genes associated with HIV/AIDS: advances in gene discovery. *Trends Genet.* **26,** 119–131 (2010).
159. Arhel, N. & Kirchhoff, F. Host proteins involved in HIV infection: new therapeutic targets. *Biochim. Biophys. Acta* **1802,** 313–321 (2010).
160. Kuritzkes, D. R. HIV-1 Entry Inhbitors: An Overview. *Curr Opin HIV AIDS* **4,** 82–87 (2009).
161. Cihlar, T. & Ray, A. S. Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Res.* **85,** 39–58 (2010).
162. Prajapati, D. G., Ramajayam, R., Yadav, M. R. & Giridhar, R. The search for potent, small molecule NNRTIs: A review. *Bioorg. Med. Chem.* **17,** 5744–5762 (2009).
163. Thierry, E., Deprez, E. & Delelis, O. Different Pathways Leading to Integrase Inhibitors Resistance. *Front Microbiol* **7,** (2017).
164. Wensing, A. M. J., van Maarseveen, N. M. & Nijhuis, M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res.* **85,** 59–74 (2010).
165. Volberding, P. A. HIV Treatment and Prevention: An Overview of Recommendations From the IAS-USA Antiretroviral Guidelines Panel. *Top Antivir Med* **25,** 17–24 (2017).
166. Mansky, L. M. HIV mutagenesis and the evolution of antiretroviral drug resistance. *Drug Resist. Updat.* **5,** 219–223 (2002).
167. Sharp, P. M. & Hahn, B. H. The evolution of HIV-1 and the origin of AIDS. *Philos Trans R Soc Lond B Biol Sci* **365,** 2487–2494 (2010).
168. Damond, F. *et al.* Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification. *AIDS Res. Hum. Retroviruses* **20,** 666–672 (2004).
169. Gao, F. *et al.* Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397,** 436–441 (1999).
170. Plantier, J.-C. *et al.* A new human immunodeficiency virus derived from gorillas. *Nat. Med.* **15,** 871–872 (2009).
171. HIV sequence database. Available at: https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html. (Accessed: 5th June 2017)
172. D'arc, M. *et al.* Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc. Natl. Acad. Sci. U.S.A.* **112,** E1343-1352 (2015).
173. Tebit, D. M. & Arts, E. J. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis* **11,** 45–56 (2011).
174. Sauter, D. *et al.* Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. *Cell Host Microbe* **6,** 409–421 (2009).
175. Mourez, T., Simon, F. & Plantier, J.-C. Non-M Variants of Human Immunodeficiency Virus Type 1. *Clin Microbiol Rev* **26,** 448–461 (2013).
176. Preston, B. D., Poiesz, B. J. & Loeb, L. A. Fidelity of HIV-1 reverse transcriptase. *Science* **242,** 1168–1171 (1988).
177. Smyth, R. P., Davenport, M. P. & Mak, J. The origin of genetic diversity in HIV-1. *Virus Res.* **169,** 415–429 (2012).
178. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271,** 1582–1586 (1996).
179. Chen, J. *et al.* High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 13535–13540 (2009).
180. Hu, W. S. & Temin, H. M. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci. U.S.A.* **87,** 1556–1560 (1990).

181. Onafuwa-Nuga, A. & Telesnitsky, A. The Remarkable Frequency of Human Immunodeficiency Virus Type 1 Genetic Recombination. *Microbiol Mol Biol Rev* **73,** 451–480 (2009).
182. Coffin, J. M. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J. Gen. Virol.* **42,** 1–26 (1979).
183. Negroni, M. & Buc, H. Copy-choice recombination by reverse transcriptases: reshuffling of genetic markers mediated by RNA chaperones. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 6385–6390 (2000).
184. Simon-Loriere, E., Rossolillo, P. & Negroni, M. RNA structures, genomic organization and selection of recombinant HIV. *RNA Biology* (2011). doi:10.4161/rna.8.2.15193
185. Zhang, J. & Temin, H. M. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. *Science* **259,** 234–238 (1993).
186. Simon-Loriere, E. & Holmes, E. C. Why do RNA viruses recombine? *Nat. Rev. Microbiol.* **9,** 617–626 (2011).
187. Ramirez, B. C., Simon-Loriere, E., Galetto, R. & Negroni, M. Implications of recombination for HIV diversity. *Virus Res.* **134,** 64–73 (2008).
188. Quiñones-Mateu, M. E., Albright, J. L., Mas, A., Soriano, V. & Arts, E. J. Analysis of pol gene heterogeneity, viral quasispecies, and drug resistance in individuals infected with group O strains of human immunodeficiency virus type 1. *J. Virol.* **72,** 9002–9015 (1998).
189. Peeters, M. *et al.* Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient. *J. Virol.* **73,** 7368–7375 (1999).
190. Takehisa, J. *et al.* Human immunodeficiency virus type 1 intergroup (M/O) recombination in cameroon. *J. Virol.* **73,** 6810–6820 (1999).
191. Yamaguchi, J. *et al.* HIV infections in northwestern Cameroon: identification of HIV type 1 group O and dual HIV type 1 group M and group O infections. *AIDS Res. Hum. Retroviruses* **20,** 944–957 (2004).
192. Fan, J., Negroni, M. & Robertson, D. L. The distribution of HIV-1 recombination breakpoints. *Infect. Genet. Evol.* **7,** 717–723 (2007).
193. Archer, J. *et al.* Identifying the important HIV-1 recombination breakpoints. *PLoS Comput. Biol.* **4,** e1000178 (2008).
194. Galli, A. *et al.* Patterns of Human Immunodeficiency Virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. *J. Virol.* **84,** 7651–7661 (2010).
195. Woo, J., Robertson, D. L. & Lovell, S. C. Constraints from protein structure and intra-molecular coevolution influence the fitness of HIV-1 recombinants. *Virology* **454,** 34–39 (2014).
196. Töpfer, A. *et al.* Probabilistic inference of viral quasispecies subject to recombination. *J. Comput. Biol.* **20,** 113–123 (2013).
197. Routh, A., Chang, M. W., Okulicz, J. F., Johnson, J. E. & Torbett, B. E. CoVaMa: Co-Variation Mapper for disequilibrium analysis of mutant loci in viral populations using next-generation sequence data. *Methods* **91,** 40–47 (2015).
198. Liao, H.-X. *et al.* Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496,** 469–476 (2013).
199. Beaumont, E. *et al.* Matrix and envelope coevolution revealed in a patient monitored since primary infection with human immunodeficiency virus type 1. *J. Virol.* **83,** 9875–9889 (2009).
200. Simon-Loriere, E. *et al.* Molecular Mechanisms of Recombination Restriction in the Envelope Gene of the Human Immunodeficiency Virus. *PLOS Pathogens* **5,** e1000418 (2009).
201. Gasser, R. *et al.* Buffering deleterious polymorphisms in highly constrained parts of HIV-1 envelope by flexible regions. *Retrovirology* **13,** 50 (2016).
202. Engelman, A., Bushman, F. D. & Craigie, R. Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.* **12,** 3269–3275 (1993).
203. Faure, A. *et al.* HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res.* **33,** 977–986 (2005).
204. Cai, M. *et al.* Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* **4,** 567–577 (1997).
205. 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).
206. Dyda, F. *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266,** 1981–1986 (1994).
207. Eijkelenboom, A. P. *et al.* The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Curr. Biol.* **7,** 739–746 (1997).

208. Zheng, R., Jenkins, T. M. & Craigie, R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 13659–13664 (1996).
209. McKee, C. J. *et al.* Dynamic modulation of HIV-1 integrase structure and function by cellular lens epithelium-derived growth factor (LEDGF) protein. *J. Biol. Chem.* **283,** 31802–31812 (2008).
210. Hare, S. *et al.* Structural basis for functional tetramerization of lentiviral integrase. *PLoS Pathog.* **5,** e1000515 (2009).
211. Lu, R., Vandegraaff, N., Cherepanov, P. & Engelman, A. Lys-34, dispensable for integrase catalysis, is required for preintegration complex function and human immunodeficiency virus type 1 replication. *J. Virol.* **79,** 12584–12591 (2005).
212. Vincent, K. A., Ellison, V., Chow, S. A. & Brown, P. O. Characterization of human immunodeficiency virus type 1 integrase expressed in Escherichia coli and analysis of variants with amino-terminal mutations. *J. Virol.* **67,** 425–437 (1993).
213. Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P. & Craigie, R. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 3428–3432 (1993).
214. Heuer, T. S. & Brown, P. O. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry* **36,** 10655–10665 (1997).
215. Esposito, D. & Craigie, R. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* **17,** 5832–5843 (1998).
216. Chen, A., Weber, I. T., Harrison, R. W. & Leis, J. Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat Ends. *J. Biol. Chem.* **281,** 4173–4182 (2006).
217. Engelman, A., Hickman, A. B. & Craigie, R. The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J. Virol.* **68,** 5911–5917 (1994).
218. Harper, A. L., Skinner, L. M., Sudol, M. & Katzman, M. Use of patient-derived human immunodeficiency virus type 1 integrases to identify a protein residue that affects target site selection. *J. Virol.* **75,** 7756–7762 (2001).
219. 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. *PNAS* **102,** 17308–17313 (2005).
220. Rahman, S., Lu, R., Vandegraaff, N., Cherepanov, P. & Engelman, A. Structure-based mutagenesis of the integrase-LEDGF/p75 interface uncouples a strict correlation between in vitro protein binding and HIV-1 fitness. *Virology* **357,** 79–90 (2007).
221. Wang, J.-Y., Ling, H., Yang, W. & Craigie, R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J* **20,** 7333–7343 (2001).
222. Berthoux, L., Sebastian, S., Muesing, M. A. & Luban, J. The role of lysine 186 in HIV-1 integrase multimerization. *Virology* **364,** 227–236 (2007).
223. Serrao, E. *et al.* A symmetric region of the HIV-1 integrase dimerization interface is essential for viral replication. *PLoS ONE* **7,** e45177 (2012).
224. 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).
225. Li, X., Krishnan, L., Cherepanov, P. & Engelman, A. Structural biology of retroviral DNA integration. *Virology* **411,** 194–205 (2011).
226. Cannon, P. M., Byles, E. D., Kingsman, S. M. & Kingsman, A. J. Conserved sequences in the carboxyl terminus of integrase that are essential for human immunodeficiency virus type 1 replication. *J. Virol.* **70,** 651–657 (1996).
227. Lutzke, R. A., Vink, C. & Plasterk, R. H. Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res* **22,** 4125–4131 (1994).
228. Zhao, Z. *et al.* Subunit-specific protein footprinting reveals significant structural rearrangements and a role for N-terminal Lys-14 of HIV-1 Integrase during viral DNA binding. *J. Biol. Chem.* **283,** 5632–5641 (2008).
229. Lutzke, R. A. & Plasterk, R. H. Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. *J. Virol.* **72,** 4841–4848 (1998).

230. Engelman, A. & Cherepanov, P. Retroviral Integrase Structure and DNA Recombination Mechanism. *Microbiology Spectrum* **2,** (2014).
231. Michel, F. *et al.* Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. *EMBO J.* **28,** 980–991 (2009).
232. Cherepanov, P. *et al.* HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* **278,** 372–381 (2003).
233. 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).
234. Johnson, B. C., Métifiot, M., Ferris, A., Pommier, Y. & Hughes, S. H. A homology model of HIV-1 integrase and analysis of mutations designed to test the model. *J. Mol. Biol.* **425,** 2133–2146 (2013).
235. Yin, Z. *et al.* Crystal structure of the Rous sarcoma virus intasome. *Nature* **530,** 362–366 (2016).
236. Ballandras-Colas, A. *et al.* Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* **530,** 358–361 (2016).
237. Ballandras-Colas, A. *et al.* A supramolecular assembly mediates lentiviral DNA integration. *Science* **355,** 93–95 (2017).
238. Delelis, O., Carayon, K., Saïb, A., Deprez, E. & Mouscadet, J.-F. Integrase and integration: biochemical activities of HIV-1 integrase. *Retrovirology* **5,** 114 (2008).
239. Pauza, C. D. Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. *Virology* **179,** 886–889 (1990).
240. Engelman, A., Mizuuchi, K. & Craigie, R. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67,** 1211–1221 (1991).
241. Bushman, F. D., Fujiwara, T. & Craigie, R. Retroviral DNA integration directed by HIV integration protein in vitro. *Science* **249,** 1555–1558 (1990).
242. Kulkosky, J. & Skalka, A. M. Molecular mechanism of retroviral DNA integration. *Pharmacol. Ther.* **61,** 185–203 (1994).
243. Craigie, R. & Bushman, F. D. HIV DNA Integration. *Cold Spring Harb Perspect Med* **2,** (2012).
244. Skalka, A. M. & Katz, R. A. Retroviral DNA integration and the DNA damage response. *Cell Death Differ.* **12 Suppl 1,** 971–978 (2005).
245. Li, L. *et al.* Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. *EMBO J* **20,** 3272–3281 (2001).
246. Yoder, K. E. & Bushman, F. D. Repair of Gaps in Retroviral DNA Integration Intermediates. *J. Virol.* **74,** 11191–11200 (2000).
247. Daniel, R., Katz, R. A. & Skalka, A. M. A role for DNA-PK in retroviral DNA integration. *Science* **284,** 644–647 (1999).
248. Bushman, F. *et al.* Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* **3,** 848–858 (2005).
249. Schröder, A. R. W. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110,** 521–529 (2002).
250. La chromatine: organisation fonctionnelle du génome. Available at : http://atlasgeneticsoncology.org/Educ/ChromatinEducFr.html. (Accessed: 13th June 2017)
251. Benleulmi, M. S. *et al.* Intasome architecture and chromatin density modulate retroviral integration into nucleosome. *Retrovirology* **12,** 13 (2015).
252. Naughtin, M. *et al.* DNA Physical Properties and Nucleosome Positions Are Major Determinants of HIV-1 Integrase Selectivity. *PLoS ONE* **10,** e0129427 (2015).
253. Ciuffi, A. *et al.* A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* **11,** 1287–1289 (2005).
254. Shun, M.-C. *et al.* Identification and Characterization of PWWP Domain Residues Critical for LEDGF/p75 Chromatin Binding and Human Immunodeficiency Virus Type 1 Infectivity. *Journal of Virology* **82,** 11555–11567 (2008).
255. Matysiak, J. *et al.* Modulation of chromatin structure by the FACT histone chaperone complex regulates HIV-1 integration. *Retrovirology* **14,** 39 (2017).
256. Lopez, A. P. *et al.* The Structure-Specific Recognition Protein 1 Associates with Lens Epithelium-Derived Growth Factor Proteins and Modulates HIV-1 Replication. *J. Mol. Biol.* **428,** 2814–2831 (2016).
257. Sowd, G. A. *et al.* A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E1054-1063 (2016).
258. Marini, B. *et al.* Nuclear architecture dictates HIV-1 integration site selection. *Nature* **521,** 227–231 (2015).

259. Lelek, M. *et al.* Chromatin organization at the nuclear pore favours HIV replication. *Nat Commun* **6,** 6483 (2015).
260. Sloan, R. D. & Wainberg, M. A. The role of unintegrated DNA in HIV infection. *Retrovirology* **8,** 52 (2011).
261. Pang, S. *et al.* High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. *Nature* **343,** 85–89 (1990).
262. Butler, S. L., Hansen, M. S. & Bushman, F. D. A quantitative assay for HIV DNA integration in vivo. *Nat. Med.* **7,** 631–634 (2001).
263. Kilzer, J. M. *et al.* Roles of host cell factors in circularization of retroviral dna. *Virology* **314,** 460–467 (2003).
264. Bukrinsky, M., Sharova, N. & Stevenson, M. Human immunodeficiency virus type 1 2-LTR circles reside in a nucleoprotein complex which is different from the preintegration complex. *J Virol* **67,** 6863–6865 (1993).
265. 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* **10,** 87 (2013).
266. De Iaco, A. *et al.* TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm. *Retrovirology* **10,** 20 (2013).
267. Engelman, A. In vivo analysis of retroviral integrase structure and function. *Adv. Virus Res.* **52,** 411–426 (1999).
268. Engelman, A., Englund, G., Orenstein, J. M., Martin, M. A. & Craigie, R. Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J Virol* **69,** 2729–2736 (1995).
269. Lu, R., Ghory, H. Z. & Engelman, A. Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase. *J. Virol.* **79,** 10356–10368 (2005).
270. Briones, M. S., Dobard, C. W. & Chow, S. A. Role of human immunodeficiency virus type 1 integrase in uncoating of the viral core. *J. Virol.* **84,** 5181–5190 (2010).
271. Wu, X. *et al.* Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions with the nucleoprotein reverse transcription complex. *J. Virol.* **73,** 2126–2135 (1999).
272. Hehl, E. A., Joshi, P., Kalpana, G. V. & Prasad, V. R. Interaction between human immunodeficiency virus type 1 reverse transcriptase and integrase proteins. *J. Virol.* **78,** 5056–5067 (2004).
273. Wilkinson, T. A. *et al.* Identifying and characterizing a functional HIV-1 reverse transcriptase-binding site on integrase. *J. Biol. Chem.* **284,** 7931–7939 (2009).
274. Tekeste, S. S. *et al.* Interaction between Reverse Transcriptase and Integrase Is Required for Reverse Transcription during HIV-1 Replication. *J. Virol.* **89,** 12058–12069 (2015).
275. Dobard, C. W., Briones, M. S. & Chow, S. A. Molecular mechanisms by which human immunodeficiency virus type 1 integrase stimulates the early steps of reverse transcription. *J. Virol.* **81,** 10037–10046 (2007).
276. Ao, Z., Fowke, K. R., Cohen, E. A. & Yao, X. Contribution of the C-terminal tri-lysine regions of human immunodeficiency virus type 1 integrase for efficient reverse transcription and viral DNA nuclear import. *Retrovirology* **2,** 62 (2005).
277. 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).
278. Devroe, E., Engelman, A. & Silver, P. A. Intracellular transport of human immunodeficiency virus type 1 integrase. *J. Cell. Sci.* **116,** 4401–4408 (2003).
279. Zaitseva, L. *et al.* HIV-1 exploits importin 7 to maximize nuclear import of its DNA genome. *Retrovirology* **6,** 11 (2009).
280. Jayappa, K. D., Ao, Z., Yang, M., Wang, J. & Yao, X. Identification of critical motifs within HIV-1 integrase required for importin α3 interaction and viral cDNA nuclear import. *J. Mol. Biol.* **410,** 847–862 (2011).
281. De Houwer, S. *et al.* The HIV-1 integrase mutant R263A/K264A is 2-fold defective for TRN-SR2 binding and viral nuclear import. *J. Biol. Chem.* **289,** 25351–25361 (2014).
282. Bukovsky, A. & Göttlinger, H. Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease. *J. Virol.* **70,** 6820–6825 (1996).
283. Quillent, C., Borman, A. M., Paulous, S., Dauguet, C. & Clavel, F. Extensive regions of pol are required for efficient human immunodeficiency virus polyprotein processing and particle maturation. *Virology* **219,** 29–36 (1996).

284. Fontana, J. *et al.* Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-Inhibited Virions: a Role for Integrase in Maturation. *J. Virol.* **89,** 9765–9780 (2015).
285. 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).
286. Van Maele, B., Busschots, K., Vandekerckhove, L., Christ, F. & Debyser, Z. Cellular co-factors of HIV-1 integration. *Trends Biochem. Sci.* **31,** 98–105 (2006).
287. Busschots, K. *et al.* The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J. Biol. Chem.* **280,** 17841–17847 (2005).
288. Busschots, K. *et al.* Identification of the LEDGF/p75 binding site in HIV-1 integrase. *J. Mol. Biol.* **365,** 1480–1492 (2007).
289. Maertens, G., Cherepanov, P., Debyser, Z., Engelborghs, Y. & Engelman, A. Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75. *J. Biol. Chem.* **279,** 33421–33429 (2004).
290. Emiliani, S. *et al.* Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *J. Biol. Chem.* **280,** 25517–25523 (2005).
291. Botbol, Y., Raghavendra, N. K., 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.* **36,** 1237–1246 (2008).
292. Segura-Totten, M. & Wilson, K. L. BAF: roles in chromatin, nuclear structure and retrovirus integration. *Trends Cell Biol.* **14,** 261–266 (2004).
293. Chen, H. & Engelman, A. The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 15270–15274 (1998).
294. Skoko, D. *et al.* Barrier-to-autointegration factor (BAF) condenses DNA by looping. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 16610–16615 (2009).
295. Farnet, C. M. & Bushman, F. D. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* **88,** 483–492 (1997).
296. Li, L. *et al.* Retroviral cDNA integration: stimulation by HMG I family proteins. *J. Virol.* **74,** 10965–10974 (2000).
297. Henderson, A., Bunce, M., Siddon, N., Reeves, R. & Tremethick, D. J. High-Mobility-Group Protein I Can Modulate Binding of Transcription Factors to the U5 Region of the Human Immunodeficiency Virus Type 1 Proviral Promoter. *J. Virol.* **74,** 10523–10534 (2000).
298. 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).
299. Morozov, A., Yung, E. & Kalpana, G. V. Structure-function analysis of integrase interactor 1/hSNF5L1 reveals differential properties of two repeat motifs present in the highly conserved region. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 1120–1125 (1998).
300. Maillot, B. *et al.* Structural and functional role of INI1 and LEDGF in the HIV-1 preintegration complex. *PLoS ONE* **8,** e60734 (2013).
301. Thierry, E., Deprez, E. & Delelis, O. Different Pathways Leading to Integrase Inhibitors Resistance. *Front Microbiol* **7,** (2017).
302. Espeseth, A. S. *et al.* HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *PNAS* **97,** 11244–11249 (2000).
303. Grobler, J. A. *et al.* Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes. *PNAS* **99,** 6661–6666 (2002).
304. Hare, S. *et al.* Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572). *Mol. Pharmacol.* **80,** 565–572 (2011).
305. Delelis, O. *et al.* Impact of Y143 HIV-1 integrase mutations on resistance to raltegravir in vitro and in vivo. *Antimicrob. Agents Chemother.* **54,** 491–501 (2010).
306. Llano, M. *et al.* An Essential Role for LEDGF/p75 in HIV Integration. *Science* **314,** 461–464 (2006).
307. Rijck, J. D. *et al.* Overexpression of the Lens Epithelium-Derived Growth Factor/p75 Integrase Binding Domain Inhibits Human Immunodeficiency Virus Replication. *J. Virol.* **80,** 11498–11509 (2006).
308. 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).
309. Jurado, K. A. & Engelman, A. Multimodal mechanism of action of allosteric HIV-1 integrase inhibitors. *Expert Rev Mol Med* **15,** e14 (2013).
310. Tintori, C. *et al.* Discovery of small molecule HIV-1 integrase dimerization inhibitors. *Bioorg. Med. Chem. Lett.* **22,** 3109–3114 (2012).

311. Feng, L., Larue, R. C., Slaughter, A., Kessl, J. J. & Kvaratskhelia, M. HIV-1 integrase multimerization as a therapeutic target. *Curr. Top. Microbiol. Immunol.* **389,** 93–119 (2015).
312. Kessl, J. J. *et al.* An allosteric mechanism for inhibiting HIV-1 integrase with a small molecule. *Mol Pharmacol* (2009). doi:10.1124/mol.109.058883
313. Myers, R. E. & Pillay, D. Analysis of natural sequence variation and covariation in human immunodeficiency virus type 1 integrase. *J. Virol.* **82,** 9228–9235 (2008).
314. Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat. Biotechnol.* **15,** 871–875 (1997).
315. Rossolillo, P., Winter, F., Simon-Loriere, E., Gallois-Montbrun, S. & Negroni, M. Retrovolution: HIV-driven evolution of cellular genes and improvement of anticancer drug activation. *PLoS Genet.* **8,** e1002904 (2012).
316. Naldini, L. *et al.* In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272,** 263–267 (1996).
317. Vozzolo, L. *et al.* Gyrase B inhibitor impairs HIV-1 replication by targeting Hsp90 and the capsid protein. *J. Biol. Chem.* **285,** 39314–39328 (2010).
318. Mohammed, K. D., Topper, M. B. & Muesing, M. A. Sequential deletion of the integrase (Gag-Pol) carboxyl terminus reveals distinct phenotypic classes of defective HIV-1. *J. Virol.* **85,** 4654–4666 (2011).
319. Mathew, S. *et al.* INI1/hSNF5-interaction defective HIV-1 IN mutants exhibit impaired particle morphology, reverse transcription and integration in vivo. *Retrovirology* **10,** 66 (2013).
320. Rihn, S. J., Hughes, J., Wilson, S. J. & Bieniasz, P. D. Uneven Genetic Robustness of HIV-1 Integrase. *Journal of Virology* **89,** 552–567 (2015).
321. Ceccherini-Silberstein, F. *et al.* Characterization and structural analysis of HIV-1 integrase conservation. *AIDS Rev* **11,** 17–29 (2009).
322. Zhu, K., Dobard, C. & Chow, S. A. Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase. *J. Virol.* **78,** 5045–5055 (2004).
323. Cribier, A. *et al.* Mutations affecting interaction of integrase with TNPO3 do not prevent HIV-1 cDNA nuclear import. *Retrovirology* **8,** 104 (2011).
324. Depatureaux, A. *et al.* HIV-1 Group O Integrase Displays Lower Enzymatic Efficiency and Higher Susceptibility to Raltegravir than HIV-1 Group M Subtype B Integrase. *Antimicrobial Agents and Chemotherapy* **58,** 7141–7150 (2014).
325. Nomaguchi, M. *et al.* Natural single-nucleotide polymorphisms in the 3' region of the HIV-1 pol gene modulate viral replication ability. *J. Virol.* **88,** 4145–4160 (2014).
326. Dar, M. J. *et al.* Biochemical and virological analysis of the 18-residue C-terminal tail of HIV-1 integrase. *Retrovirology* **6,** 94 (2009).
327. Wiskerchen, M. & Muesing, M. A. Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. *J Virol* **69,** 376–386 (1995).
328. Rihn, S. J. *et al.* Extreme Genetic Fragility of the HIV-1 Capsid. *PLoS Pathog* **9,** (2013).
329. Santos, C. D. S., Tartour, K. & Cimarelli, A. A Novel Entry/Uncoating Assay Reveals the Presence of at Least Two Species of Viral Capsids During Synchronized HIV-1 Infection. *PLOS Pathogens* **12,** e1005897 (2016).

Annexes

Figure annexe 1 : Alignement de séquences des IN. Trois séquences d'intégrases ont été alignées : la première est celle de la souche de référence de laboratoire HXB2 provenant du sous-type B du groupe M, notée M-IN Hxb2 ; la deuxième est celle de l'isolat primaire du sous-type A du groupe M utilisé dans cette étude, notée M-IN A2 ; la troisième est celle de l'isolat RBF206 du groupe O utilisé dans cette étude, notée O – IN RBF206. Les résidus qui diffèrent entre les séquences sont surlignés en rose. Le taux de conservation est représenté schématique par des histogrammes en cyan.

Marine KANJA

Coévolution dans le gène pol du VIH-1 : un carrefour aux frontières de nouvelles espèces du VIH

Résumé

L'intégrase (IN) est l'une des enzymes virales assurant la réplication du VIH. La fonctionnalité des protéines qui, comme celles du VIH, ont une variabilité de séquence repose sur des résidus non conservés, en plus des acides aminés conservés entre souches, qui ont un rôle important notamment lorsqu'ils font partie de réseaux de coévolution. Ces réseaux peuvent contrecarrer l'effet délétère d'une mutation par l'introduction de mutations compensatoires ailleurs dans la protéine.

Ce travail a mis en évidence, par une étude comparative de différentes souches du VIH, des réseaux de coévolution étendus dans l'IN. Un résultat majeur est l'identification d'un nouveau motif assurant de multiples rôles dans le cycle infectieux. Le motif diffère entre les groupes M et O du VIH, mais est strictement conservé au sein de ces deux groupes en dépit d'une certaine flexibilité génétique en culture de cellules. Ceci suggère que ces groupes ont suivi des chemins évolutifs convergents bien que distincts.

Mots-clés : VIH, intégrase, réseaux de coévolution, fonctionnalité

Abstract

Integrase (IN) is one of the viral enzymes ensuring HIV replication. The functionality of proteins, which, like those from HIV, have sequence variability, relies on non-conserved residues, in addition to the conserved amino acids between strains, which have an important role especially when they are part of coevolution networks. These networks can counteract the deleterious effect of a mutation by introducing compensatory mutations elsewhere in the protein.

This work has demonstrated, through a comparative study of different strains of HIV, extensive coevolution networks in IN. A major result is the identification of a new motif that provides multiple roles in the infectious cycle. The pattern differs between HIV groups M and O, but is strictly conserved within these two groups despite some genetic flexibility in cell culture. This suggests that these groups followed convergent, although distinct, evolution pathways.

Keywords: HIV, integrase, coevolution networks, functionality