



**HAL**  
open science

# Étude des paramètres structuraux et cinétiques caractérisant les interactions intégrases rétrovirales / ADN et l'étape de 3'-processing

Kévin Carayon

► **To cite this version:**

Kévin Carayon. Étude des paramètres structuraux et cinétiques caractérisant les interactions intégrases rétrovirales / ADN et l'étape de 3'-processing. Biochimie [q-bio.BM]. École normale supérieure de Cachan - ENS Cachan, 2008. Français. NNT: . tel-00363674

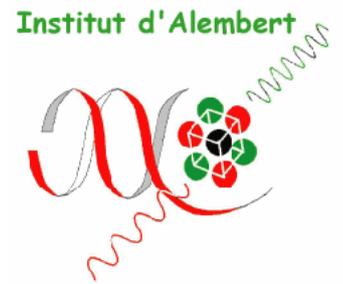
**HAL Id: tel-00363674**

**<https://theses.hal.science/tel-00363674v1>**

Submitted on 24 Feb 2009

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

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



**THESE DE DOCTORAT  
DE L'ECOLE NORMALE SUPERIEURE DE CACHAN**

Présentée par  
Monsieur Kévin CARAYON

**Pour obtenir le grade de  
DOCTEUR DE L'ECOLE NORMALE SUPERIEURE DE CACHAN**

Domaine :  
**SCIENCES DE LA VIE ET DE LA SANTE**

Sujet de la thèse :

**Etude des paramètres structuraux et cinétiques caractérisant les  
interactions intégrases rétrovirales / ADN et l'étape de 3'-processing.**

Thèse présentée et soutenue à Cachan le 04 décembre 2008 devant le jury composé de :

Jean-François Mouscadet	Directeur de Recherche, CNRS	Président
Gilles Divita	Directeur de Recherche, CNRS	Rapporteur
Corinne Ronfort	Chargée de Recherche, INRA	Rapporteur
Eric Deprez	Chargé de Recherche, CNRS	Directeur de thèse

*Laboratoire de Biotechnologie et de Pharmacologie Appliquée  
(CNRS UMR 8113)  
61, avenue du Président Wilson, 94235 Cachan CEDEX (France)*



# **REMERCIEMENTS**

Cette thèse a été effectuée au Laboratoire de Biotechnologie et de Pharmacologie génétique Appliquée, UMR 8113 du CNRS, dirigé successivement au cours de ces trois années par Christian AUCLAIR puis par Jean-François MOUSCADET que je remercie tous les deux pour m'avoir accueilli au sein de leur unité de recherche.

Je souhaite également exprimer toute ma gratitude envers les membres de mon jury de thèse de m'avoir fait l'honneur d'évaluer ce travail.

Je tiens également à témoigner toute ma reconnaissance à Jean-Claude BROCHON pour son accueil chaleureux au sein de son équipe de Biophotonique. Je tenais aussi à le remercier pour les conditions de travail idéales dans lesquelles s'est déroulée ma thèse.

Je tiens également à remercier tous les membres passés et présents de cette équipe qui m'ont soutenu tout au long de cette thèse. Merci à Patrick, Etienne et Rose. Un grand merci à Elvire pour son soutien durant mon M2 et ce début de thèse qui fut parfois bien difficile. Toutes nos discussions m'ont beaucoup apporté.

Un énorme merci à Françoise. Merci pour TOUT : ta patience lorsque je m'étais un peu sur toutes les paillasses, ton aide technique, ton humour, ta gentillesse, ton amitié, tes conseils et ton soutien de chaque instant lors des moments difficiles. Tu vas vraiment me manquer lorsque je retournerai dans mon sud natal.

Enfin, je remercie infiniment Eric pour son encadrement qui correspond à celui dont je pouvais rêver pour ma thèse. Tu m'as appris à canaliser mon énergie, gérer mon temps, à prendre le recul nécessaire sur les manip pour en tirer le maximum d'informations tout en ayant conscience de leurs limites. Merci pour toutes les discussions que nous avons pu avoir et pour tous les conseils avisés que tu as pu me donner durant toutes ces années de collaboration.

Je remercie également tous les membres du LBPA pour tous les moments que j'ai pu avoir avec eux, plus ou moins explosifs pour certains, mais j'ai passé des instants inoubliables en votre compagnie.

Je souhaite remercier en particulier Anne-Marie et Janine pour leur aide lors de mes commandes ou de mes petits problèmes administratifs du dernier moment et surtout pour tous ces moments de détente que nous avons partagés. Merci à toutes les deux.

Merci également à Martine pour les cafés quotidiens, essentiels pour garder le rythme ainsi que pour les autoclavages de LB de dernière minute hyper urgents. Merci pour ta disponibilité et ta gentillesse.

Merci à mes collègues des pièces 104 et 105 : Marie-Christine, Bernie, Fred, Julie et Séverine. Merci doublement : d'abord pour toutes les discussions scientifiques et autres, ainsi que pour les bons moments que nous avons pu partager depuis maintenant un an, mais aussi pour votre tolérance auditive qui fut mise à rude épreuve parfois je le reconnais. Tout le monde n'aurait pas accepté ou supporté d'avoir de l'opéra en guise de fond sonore à longueur de journée... Un merci tout particulier à Laurent : tes conseils, ton soutien, ton aide et ton amitié lors de cette dernière année de thèse me furent vraiment précieux.

Je remercie également Barbara et Alexandra pour leur écoute, leurs conseils personnels comme techniques (gels, chromatographies laborieuses, constructions récalcitrantes,...), la lecture de ce manuscrit et les diverses corrections de mes présentations PowerPoint. Merci à toutes les deux pour le temps que vous m'avez consacré.

Je remercie Olivier, Gladys, Aurélie, Stéphanie et tous les autres membres de l'équipe de Jean-François Mouscadet pour leur aide et leurs conseils. Merci également à Françoise pour ton soutien et tes conseils en biologie moléculaire. Enfin, merci à Géraldine pour les moments d'évasion et toutes nos discussions florales.

Durant ces trois ans, j'ai eu le plaisir de réaliser mon monitorat au sein du département Biochimie-Génie Biologique dirigé par Christian AUCLAIR secondé par Jean-François LEFLOHIC puis par Gérald PEYROCHE. Je tiens à les remercier tous les deux pour toute l'aide et la compréhension dont ils ont fait preuve durant ces trois ans afin de me permettre de gérer au mieux mes heures d'enseignement en fonction de mes plannings de manips.

Je tiens aussi à remercier tous les enseignants, techniciens et secrétaire du département BGB ainsi que tous les étudiants que j'ai eu l'occasion d'encadrer pour tous les moments agréables que j'ai passé avec eux. Qui aurait cru que je serai amené un jour à encadrer avec autant de conviction les fameux TP « rats » de la prépa agreg... pas moi ! Et pourtant, ils vont

me manquer ces TP. J'en profite pour te remercier Catherine pour m'avoir d'une part mis en relation avec Eric et d'autre part pour m'avoir transmis tout ce dont j'ai pu avoir besoin pour encadrer ces TP.

Merci pour toute la confiance dont vous avez fait preuve au cours de ces trois ans.

Enfin merci à tous mes amis (Nico, Marine, Olwenn, Florent, Carine, Georges, Laurence et Christelle) pour leur présence et leur soutien de chaque instant ainsi qu'aux rockeurs et rockeuses de l'ENS qui m'ont permis, parfois un peu à leur dépend, d'évacuer mon trop plein d'énergie et de stress. Un petit mot pour souhaiter la bienvenue au petit Maël qui s'est fait autant désirer que cette thèse. Ces deux aboutissements quasi-simultanés ne peuvent qu'être de bon augure pour nos deux avenir.

Je remercie toute ma famille et belle-famille pour leur soutien. Je veux remercier tout particulièrement mes parents pour leur confiance qu'ils m'ont témoigné durant toutes ces années lors de tous les choix que j'ai eu à faire. Si j'en suis arrivé là, c'est grâce à vous deux : merci pour tout.

Mon dernier merci, et le plus important, est pour toi Marjo : tu es à mes cotés depuis bientôt 10 ans. Ton amour, ton soutien et ta compréhension m'ont toujours accompagné depuis ce 16 janvier 1999. Tout n'a pas été facile. Ca a même été très dur parfois, mais tous les combats que nous avons eu à mener (passés comme présents), nous ont beaucoup appris sur nous même comme sur les autres et au final toutes ces difficultés rencontrées nous ont toujours rapprochés et renforcés. J'ai confiance en notre bonne étoile et je suis convaincu que le plus beau reste à venir. Il faut juste y croire mon cœur.

Merci à tous.



# TABLE DES MATIERES

REMERCIEMENTS .....	1
TABLE DES MATIERES .....	5
TABLE DES FIGURES .....	7
ABREVIATIONS .....	8
PREAMBULE .....	11
INTRODUCTION .....	15
I. LA FAMILLE DES RETROVIRIDAE (00.061) .....	17
I.A. <i>La structure taxonomique de la famille des Retroviridae.</i> .....	19
I.B. <i>Le genre Lentivirus</i> .....	20
<i>Le VIH-1</i> .....	21
I.B.1. Présentation, historique et origine du VIH-1 .....	21
I.B.2. Pathogénicité du VIH-1 .....	23
I.B.3. Cycle réplicatif du VIH-1 .....	27
I.B.3.a. Structure et composition du virion .....	27
I.B.3.b. Les étapes précoces du cycle de réplication .....	29
I.B.3.c. L'intégration .....	35
I.B.3.d. Les étapes tardives du cycle de réplication .....	35
I.C. <i>Le genre Spumavirus</i> .....	41
I.D. <i>Les traitements anti-rétroviraux actuels.</i> .....	43
II. L'INTEGRASE DU VIH, SA STRUCTURE, SES FONCTIONS .....	47
II.A. <i>Les activités catalytiques de l'intégrase du VIH-1</i> .....	47
II.A.1. Les activités de l'intégrase in-vivo .....	47
II.A.2. Les activités de l'intégrase observées in vitro .....	53
II.A.3. Modélisation du mécanisme catalytique de l'intégrase du VIH-1 .....	55
II.A.3.a. Les réactions catalysées par l'IN sont de type $S_N2$ .....	57
II.A.3.b. Dissection de la catalyse à l'échelle atomique .....	59
II.B. <i>Structures et fonctions des trois domaines de l'intégrase</i> .....	65
II.B.1. Le domaine N-Terminal .....	65
II.B.2. Le Core Catalytique .....	69
II.B.3. Le domaine C-Terminal .....	73
II.C. <i>Structure de l'intégrase : des doubles domaines à la modélisation</i> .....	75
II.C.1. Structure des doubles domaines .....	75
II.C.1.a. Structure du double domaine core catalytique / C-terminal .....	75
II.C.1.b. Structure du double domaine N-terminal / Core catalytique .....	79
II.C.2. Modélisation de la structure de l'intégrase entière .....	81
II.D. <i>Les inhibiteurs de l'intégrase du VIH-1</i> .....	89
II.D.1. Les inhibiteurs de la fixation de l'ADN viral dit « InBI » .....	89
II.D.2. Les inhibiteurs du transfert de brin dit « InSTI » .....	91
II.D.3. Les autres classes d'inhibiteurs de l'intégrase .....	93
PRESENTATION DE LA THESE ET OBJECTIFS .....	95
MATERIEL ET METHODES .....	98
I. PROTOCOLE DE PURIFICATION DE L'INTEGRASE ET TESTS D'ACTIVITES IN VITRO .....	99
I.A. <i>Clonage et mutagenèse de l'intégrase</i> .....	99
I.A.1. Mutations ponctuelles sur l'intégrase du VIH-1 .....	99
I.A.2. Délétions sur l'IN du VIH-1 .....	100
I.B. <i>Expression</i> .....	102
I.C. <i>Protocole de purification</i> .....	103
I.D. <i>Marquage enzymatique de l'IN avec un petit fluorophore</i> .....	105
I.E. <i>Tests d'activités</i> .....	106
I.E.1. Test utilisant l'anisotropie de fluorescence statique .....	106
I.E.1.a. Présentation de l'appareillage : le Beacon et l'Eclipse .....	106
I.E.1.b. Test d'activité .....	107

I.E.2. Test d'activité utilisant les oligonucléotides radiomarqués.....	108
II. PRESENTATION DU PRINCIPE DE LA FLUORESCENCE ET DES TECHNIQUES UTILISEES .....	110
II.A. <i>Le phénomène de fluorescence</i> .....	110
II.B. <i>L'anisotropie de fluorescence</i> .....	112
II.B.1. Principe et anisotropie statique.....	112
II.B.2. L'anisotropie résolue en temps.....	114
II.B.3. Description du montage utilisé.....	116
II.C. <i>Le Transfert d'Energie Résonant de Fluorescence (FRET)</i> .....	118
II.D. <i>La Spectroscopie de Fluorescence Corrélée (FCS)</i> .....	120
II.D.1. Fluorescence sous excitation biphotonique .....	120
II.D.2. Principe de la Spectroscopie de Fluorescence Corrélée (FCS).....	122
II.D.3. Description du montage utilisé.....	125
<b>RESULTATS</b> .....	<b>127</b>
I. RELATION ENTRE L'ETAT OLIGOMERIQUE DE L'IN DU VIH-1 SUR L'ADN ET SON ACTIVITE ENZYMATIQUE	128
I.A. <i>Introduction</i> .....	128
I.B. <i>Manuscrit</i> .....	129
I.C. <i>Conclusion</i> .....	143
II. ETUDE DU MECANISME DE RECONNAISSANCE IN / ADN : UN MODE D'INTERACTION A L'ADN SPECIFIQUE MIS EN EVIDENCE PAR UNE IN MARQUEE AVEC UNE TECHNIQUE ENZYMATIQUE .....	145
II.A. <i>Introduction</i> .....	145
II.B. <i>Manuscrit</i> .....	146
II.C. <i>Conclusion</i> .....	159
III. ETUDE DIFFERENTIELLE SELON LE CONTEXTE CATIONIQUE AU SITE ACTIF DE L'IMPLICATION DU DOMAINE N-TERMINAL DE L'IN DU VIH-1 LORS DU PROCESSUS MULTIMERIQUE DE FIXATION A L'ADN VIRAL .....	161
III.A. <i>Introduction</i> .....	161
III.B. <i>Matériel et méthodes</i> .....	163
III.C. <i>Résultats</i> .....	166
III.D. <i>Discussion</i> .....	169
III.E. <i>Figures et Tableaux</i> .....	172
<b>CONCLUSION ET DISCUSSION GENERALE</b> .....	<b>177</b>
<b>ANNEXES</b> .....	<b>190</b>
ANNEXE 1 : COMPOSITION DES TAMPONS UTILISES LORS DE LA PURIFICATION DE L'IN.....	191
ANNEXE 2 : INTEGRASE ET INTEGRATION, ACTIVITES BIOCHIMIQUES DE L'INTEGRASE DU VIH-1 .....	192
<i>Résumé</i> .....	192
<i>Manuscrit</i> .....	192
<b>BIBLIOGRAPHIE</b> .....	<b>223</b>

# TABLE DES FIGURES

- FIGURE 1 : ORGANISATION GENOMIQUE D'UN RETROVIRUS DIT « SIMPLE »  
FIGURE 2 : ARBRE PHYLOGENETIQUE DES RETROVIRUS  
FIGURE 3 : ARBRE PHYLOGENETIQUE DES VIH-1 ET -2  
FIGURE 4 : CINETIQUE DE L'INFECTION AU VIH  
FIGURE 5 : CYCLE DE REPLICATION DU VIH-1  
FIGURE 6 : STRUCTURE DU VIRION MATURE DU VIH-1  
FIGURE 7 : PROCESSUS D'ENTREE DU VIH-1 DANS LA CELLULE  
FIGURE 8 : ETAPES PRECOCES DU CYCLE DE REPLICATION DU VIH  
FIGURE 9 : TRANSCRIPTION INVERSE REALISEE PAR LA RT  
FIGURE 10 : REGULATION DE LA TRANSCRIPTION POST-INTEGRATIVE DU VIH  
FIGURE 11 : EPISSAGE ALTERNATIF DE L'ARNM DU VIH-1  
FIGURE 12 : INFLUENCE DE NEF SUR L'ENVIRONNEMENT CELLULAIRE  
FIGURE 13 : MODIFICATIONS POST-TRADUCTIONNELLES ET ADRESSAGE A LA MEMBRANE  
FIGURE 14 : MATURATION DU VIRION DU VIH PAR LA PROTEASE  
FIGURE 15 : ORGANISATION GENOMIQUE DES SPUMAVIRUS.  
FIGURE 16 : MECANISME D'INTERACTION METABOLIQUE DROGUE / DROGUE  
FIGURE 17 : L'INTEGRATION DE L'ADN DU VIH-1 IN VIVO  
FIGURE 18 : REACTIONS CATALYSEES PAR L'IN IN VIVO  
FIGURE 19 : REACTIONS CATALYSEES PAR L'IN IN VITRO  
FIGURE 20 : CATION DIVALENT ET INTERACTION IN/ADN ET IN/DROGUES  
FIGURE 21 : MECANISME CATALYTIQUE DE L'IN DE TYPE  $S_N2$   
FIGURE 22 : ACTIVITE DE L'IN EN FONCTION DU PH ET DU CATION DIVALENT  
FIGURE 23 : DIAGRAMME ENERGETIQUE DE LA REACTION DE  $S_N2$  CATALYSEE PAR L'IN  
FIGURE 24 : VUE EN STEREO DES 2 SITES DE FIXATION DES CATIONS DIVALENTS DE L'IN D'ASV  
FIGURE 25 : COMPARAISON DES SITES ACTIFS DE L'IN D'ASV, LA RNASE H ET DU FRAGMENT DE KLENOW  
FIGURE 26 : CONFIGURATION DU SITE ACTIF DE L'IN DU VIH-1 AVEC 2  $Mg^{2+}$  AU SITE ACTIF.  
FIGURE 27 : STRUCTURE ET FONCTION DES 3 DOMAINES DE L'IN  
FIGURE 28 : STRUCTURES DU DOMAINE N-TER DE L'IN DU VIH-1  
FIGURE 29 : STRUCTURE DU DIMERE DU DOMAINE N-TER (FORME D)  
FIGURE 30 : STRUCTURE DU MONOMERE DU CC DE L'IN DU VIH-1  
FIGURE 31 : LA BOUCLE FLEXIBLE DU CC : CONSERVATION ET MOUVEMENTS.  
FIGURE 32 : STRUCTURE DU DIMERE DU CC, EN PRESENCE DE  $Mg^{2+}$   
FIGURE 33 : STRUCTURE DU DIMERE DU DOMAINE C-TER  
FIGURE 34 : REGIONALISATION DU DOMAINE C-TER DE L'IN DU VIH-1  
FIGURE 35 : STRUCTURE DU DOUBLE DOMAINE CC/C-TER DE L'IN  
FIGURE 36 : COMPARAISON DES DOUBLES DOMAINE CC/C-TER DES IN DU VIH-1, RSV ET VIS.  
FIGURE 37 : STRUCTURE DU DOUBLE DOMAINE N-TER/CC DE L'IN  
FIGURE 38 : POSITIONNEMENT DU PHOSPHATE AU NIVEAU DU SITE ACTIF  
FIGURE 39 : MODELISATION DE LA STRUCTURE DE L'IN A PARTIR DES DOUBLES DOMAINES.  
FIGURE 40 : MODELE 1 DU TETRAMERE D'IN, COMPARAISON AVEC LE COMPLEXE TRANSPOSASE  
TN5/ADN  
FIGURE 41 : MODELE 2 DU TETRAMERE D'IN, COMPLEXE AVEC LES ADN VIRAUX ET CIBLE  
FIGURE 42 : COMPARAISON DES SURFACES DU COMPLEXE 3D IN/ADN ET DES 2 MODELES CITES  
FIGURE 43 : LES STYRYLQUINOLINES : INHIBITEURS DE TYPE INBI  
FIGURE 44 : LES INHIBITEURS DU TRANSFERT DE BRIN OU INSTI  
FIGURE 45 : DIAGRAMME DE JABLONSKI  
FIGURE 46 : SUPERPOSITION DU SPECTRE D'EXCITATION ET D'EMISSION D'UN FLUOROPHORE.  
FIGURE 47 : PRINCIPE DE L'ANISOTROPIE DE FLUORESCENCE  
FIGURE 48 : EXEMPLE DE DECLINS D'ANISOTROPIE  
FIGURE 49 : MONTAGE UTILISE POUR LES MESURES DE TFA  
FIGURE 50 : LA FLUORESCENCE SOUS EXCITATION BI-PHOTONIQUE  
FIGURE 52 : MONTAGE UTILISE POUR LES MESURES DE FCS  
FIGURE 53 : MODELE DE LA FORMATION DU COMPLEXE DIMERE D'IN / ADN COMPETENT POUR LE 3'-  
PROCESSING  
FIGURE 54 : MODELE DU DIMERE D'IN EN PRESENCE D'ADN VIRAL PROCESSE

## ABBREVIATIONS

5CITEP	(Z)-3-(5-chloro-1H-indol-3-yl)-3-hydroxy-1-(2H-tetrazol-5-yl) prop-2-en-1-one
ASV	« Avian Sarcoma Virus »
BAF	« Barrier-to-Autointegration Factor »
Bel	« Between Env and LTR »
BIV	« Bovine Immunodeficiency Virus »
CA	Protéine de capsid
CC	Core Catalytique
CCR5	« Chemokine (C-C motif) Receptor 5 »
CPI	Complexe de Préintégration
C-ter	Domaine C-terminal
CXCR4	« Chemokine (C-X-C motif) Receptor 4 »
DIBA-1	N-(4-sulfamoylphenyl)-2-[2[(4sulfamoylphenyl)carbamoyl]phenyl]disulfanyl benzamide
env	« envelope »
FCS	« Fluorescence Correlation spectroscopy »
FDA	« Food and Drug Administration »
FIV	« Feline Immunodeficiency Virus »
FRET	« Fluorescence Resonance Energy Transfer »
gag	« group-specific antigen »
HAART	« Highly Active Antiretroviral Therapy »
HSP60	« Heat Shock Protein 60 »
IC <sub>50</sub>	Concentration de drogue nécessaire pour observer 50% d'inhibition
IN	Intégrase
InBI	« Integrase Binding Inhibitors»
INI1	« Integrase interactor-1 »
InSTI	« Integrase Strand Transfert Inhibitor »
Kd	Constante de dissociation
LEDGF/p75	« Lens Epithelium-derived Growth Factor »
LT CD4+	Lymphocytes T CD4+
LTR	« Long Terminal Repeat »

MA	Protéine de matrice
NC	Protéine de nucléocapside
Nef	« Negative factor »
NF- $\kappa$ B	« Nuclear factor-kappa beta »
NLS	« Nuclear Localization Sequence »
N-ter	Domaine N-terminal
ORF	Open Reading Frames
PCR	« Polymerase Chain Reaction »
PDB	« Protein Databank Brookhaven »
PFV	« Primate Foamy Virus »
pol	« polymerase »
PR	Protease
R	« Redundant » séquence redondante située aux extrémités de l'ADN <sub>viral</sub>
Rev	« Regulator of Viral Expression »
RMN	Résonance Magnétique Nucléaire
RSV	« Rous Sarcoma Virus »
RT	Transcriptase inverse (Reverse)
SIDA	Syndrome de l'Immunodéficience Acquise
SUgp120	Sous-Unité glycoprotéine de 120 kDa
TAMRA	carboxy tétraméthyl rhodamine
TAR	« Trans-acting responsive element »
Tat	Trans-Activator for viral Transcription
TFA	Anisotropie de Fluorescence résolue en Temps
TMgp41	glycoprotéine transmembranaire de 41 kDa
U3	« Unique sequence element at the 3'end of the viral DNA »
U5	« Unique sequence element at the 5'end of the viral DNA »
Vif	« Virion Infectivity Factor »
VIH	Virus de l'Immunodéficience Humain
VIS	Virus de l'Immunodéficience Simien
Vpr	« Viral Protein Regulatory »
Vpu	« Viral Protein Unknown »



# **PREAMBULE**

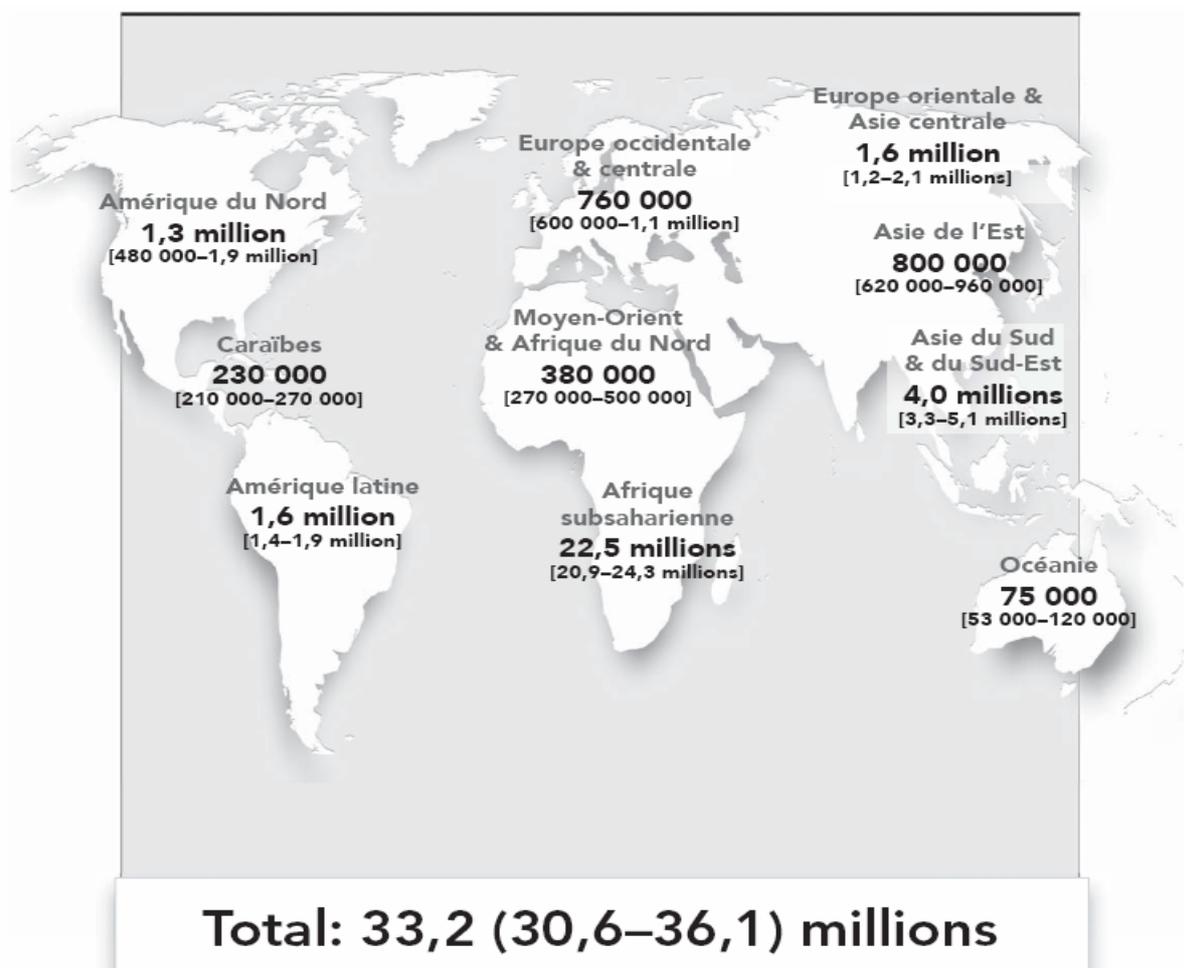
Les dernières données sur le SIDA communiquées conjointement par l'ONUSIDA et l'Organisation Mondiale de la Santé (OMS) montrent que, chaque jour dans le monde, plus de 6800 personnes sont infectées par le Virus de l'Immunodéficience Humaine (VIH) et que plus de 5700 en meurent. Les raisons majeures de cette évolution sont le manque d'accès aux services de prévention et de traitement de l'infection.

Les chiffres au niveau mondial parlent d'eux-mêmes :

Personnes vivant avec le SIDA en 2007	33,2 millions [30,6 - 36,1 millions]
Nouveaux cas d'infection au VIH en 2007	2,5 millions [1,8 - 4,1 millions]
Décès du au SIDA en 2007	2,1 millions [1,9 - 2,4 millions]

Au niveau géographique :

### Estimation du nombre d'adultes et d'enfants atteints du VIH en 2007



La pandémie du VIH reste le défi infectieux le plus grave en matière de santé publique. Actuellement, le SIDA est à lui seul responsable de la diminution de l'espérance de vie observée dans plusieurs pays particulièrement touchés par cette infection. En octobre

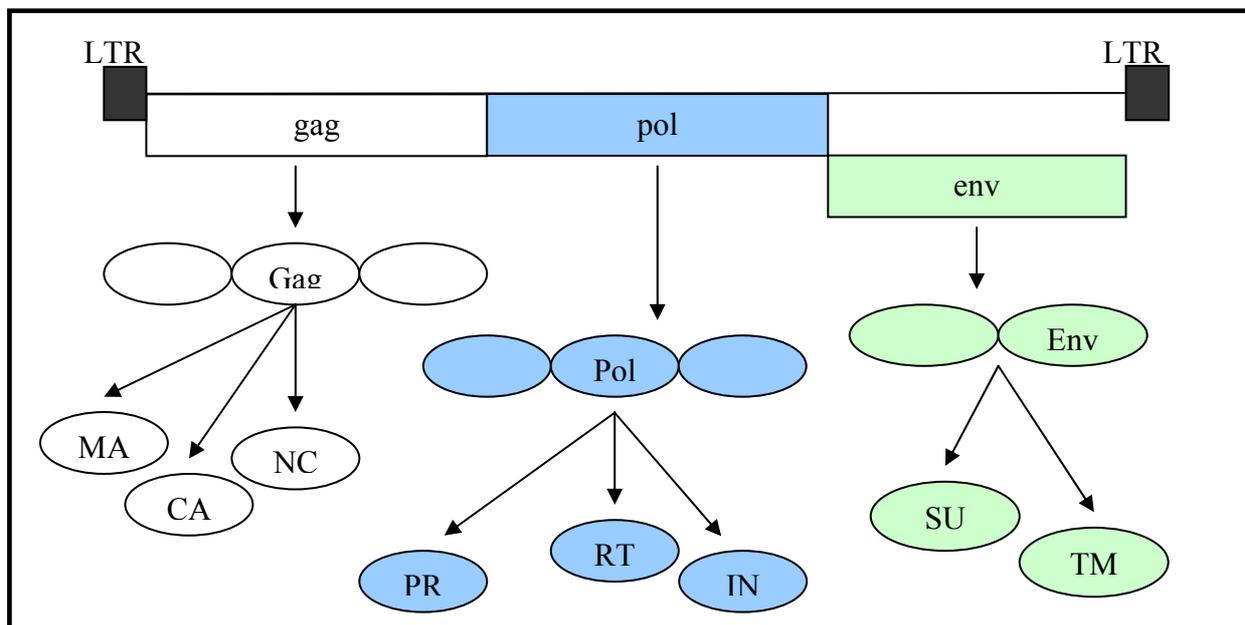
2002, l'OMS a fait un communiqué : « On estime actuellement à 47 ans l'espérance de vie à la naissance en Afrique subsaharienne. Sans le SIDA, elle serait d'environ 62 ans ». Un exemple marquant est la diminution observée en Afrique du Sud où l'espérance de vie est passée de 62,8 ans en 1990 à 51,3 ans en 2006 alors que parallèlement, en France, l'espérance de vie est passée de 77,4 à 80,7 ans (estimations OMS).

A l'heure actuelle, la recherche médicale a beaucoup avancé et plusieurs médicaments à activité anti-rétrovirale sont disponibles sur le marché. Cependant, cette avancée est contre balancée par deux phénomènes. Le premier est un accès difficile des populations pauvres à ces traitements coûteux et le second est l'apparition de virus multirésistants aux différents traitements prescrits. En effet, le virus du VIH se caractérise par une très grande variabilité génétique expliquée par son mode de répllication. A chaque fois qu'il est mis en présence d'un nouveau composé à activité anti-rétrovirale, des mutations de résistances, qui sont propres à chaque molécule, ou famille de molécules, apparaissent permettant ainsi au virus de se multiplier efficacement à nouveau. Il est donc important de comprendre de manière précise pour chaque étape du cycle de multiplication rétroviral potentiellement ciblée par des inhibiteurs le mécanisme catalytique et les différentes interactions qui y sont associées. Ceci peut permettre de trouver, pour une même cible, différentes classes d'inhibiteurs qui posséderont des spectres de mutations de résistances non recouvrant, de manière à limiter au maximum les phénomènes de résistance croisée.

De ce fait, la recherche de nouvelles cibles thérapeutiques représente un enjeu de taille dans la lutte contre ce fléau. Si notre étude ne s'inscrit pas directement dans la recherche de nouvelles molécules à destinée thérapeutique, elle a pour but de comprendre le déroulement mécanistique d'une étape clé du cycle de multiplication du VIH : l'intégration de son génome au sein de celui de la cellule hôte. Cette étape est catalysée par une enzyme codée par le génome viral : l'intégrase (IN). Depuis le début des années 90, beaucoup de molécules montrant une activité inhibitrice de l'IN ont été identifiées à l'aide de tests menés *in vitro*. Cependant, que se soit lors du passage à des tests d'inhibition *ex vivo* ou durant les premières phases des tests cliniques, la majorité des molécules s'est révélée ou inefficace contre la multiplication virale, ou trop toxique pour être développée dans le cadre des multithérapies. Après plus de 15 ans d'études intensives, un seul inhibiteur de l'IN, le Raltegravir, a été accepté par la Food and Drug Administration (FDA) en novembre 2007.



# **INTRODUCTION**



**Figure 1 : Organisation génomique d'un rétrovirus dit « simple »**

Organisation génomique typique des rétrovirus dit « simples » avec gag-pol-env et les protéines virales issues de la protéolyse des trois polyprotéines traduites.

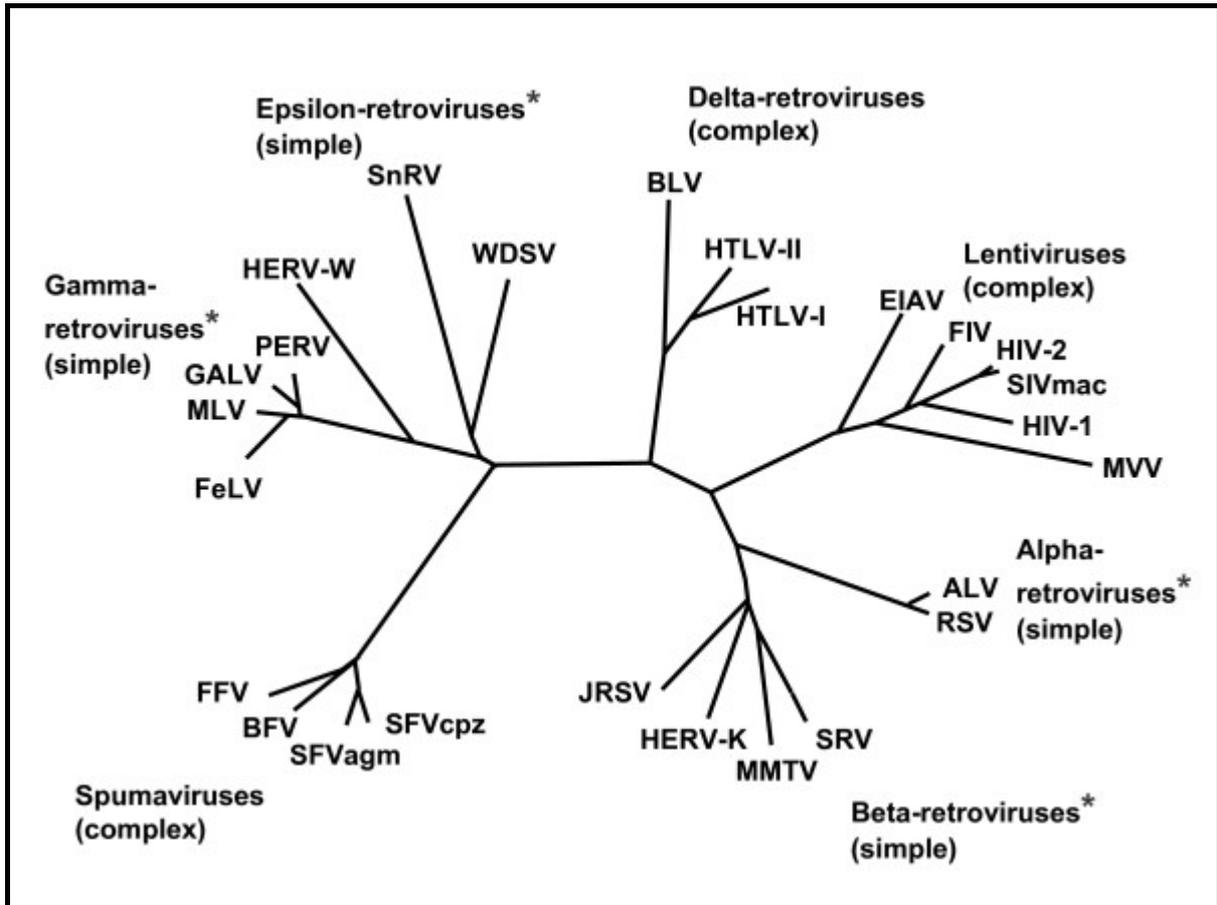
## ***I. La Famille des Retroviridae (00.061)***

C'est une large famille de virus infectant essentiellement les vertébrés. Ces virus sont associés à de nombreuses maladies qui sont le plus souvent dite à évolution lente comme des maladies dégénératives du système nerveux central et du système immunitaire. Ils peuvent également être responsables d'infections chroniques avec réplication continue du virus, mais sans pathologie détectable. Malgré la diversité des espèces infectées, et des interactions avec l'hôte, tous les rétrovirus connus ont une structure et un mode de réplication similaire. Leur matériel génétique se compose d'ARN de polarité positive. Jusqu'à présent, deux familles de virus, les *retroviridae* et les *hepadnavirus*, possèdent une enzyme permettant la transcription inverse d'ARN simple brin en ADN bicaténaire : la transcriptase inverse (RT). Le génome des rétrovirus code en plus pour une enzyme qui va intégrer l'ADN proviral néosynthétisé par la transcription inverse au sein du génome de la cellule hôte : l'intégrase (IN).

Les rétrovirus se composent d'une double copie de leur génome sous forme d'ARN de polarité positive (7 à 12 kb), protégée par la nucléocapside de nature protéique. L'ensemble est lui-même inclus dans la capsid, elle aussi de nature protéique à géométrie sphérique ou conique. Ils présentent tous une enveloppe constituée d'une bicouche phospholipidique sphérique d'environ 100nm de diamètre, composée de protéines virales mais aussi de différentes protéines membranaires. Certaines protéines, de part leur structure et leur affinité pour des protéines membranaires de l'hôte, vont donner le tropisme du rétrovirus pour un type d'hôte (famille, espèces, ...) et un type cellulaire (cellules du système immunitaire, du système nerveux, ...).

Les gènes codent pour des polyprotéines qui après protéolyse donneront plusieurs protéines structurales. Ils sont dans un ordre invariable : gag – pol – env (cf. figure 1). gag donne les protéines de la matrice (MA), de la capsid (CA) et de la nucléocapsid (NC). pol code pour les trois enzymes virales que sont la protéase (PR), la transcriptase inverse (RT) et l'intégrase (IN). Enfin, env code pour les protéines d'enveloppe. Lorsque le génome du rétrovirus ne code que pour Gag, Pol et Env, il est dit « simple ». Dans d'autres cas, le génome peut coder pour des protéines de régulation dites accessoires, le rétrovirus est alors dit « complexe ».

Lors de la phase précoce de la réplication, l'ARN viral est converti en ADN double brin par la RT. Dans un deuxième temps, l'IN virale l'intègre au sein de l'ADN de la cellule



*Figure 2 : Arbre phylogénétique des rétrovirus*

Arbre des six genres appartenant aux *orthoretrovirinae*. A noter le positionnement des *spumaretrovirinae*. D'après (Weiss, 2006)

infectée. L'intégration faite, le provirus peut s'exprimer ou non, mais il restera toujours présent de manière latente dans la cellule. Cette latence permet au rétrovirus d'échapper au système immunitaire et de ressurgir lorsque sa vigilance est amoindrie, comme dans les cas de stress intense ou d'immunodépression.

## I.A. La structure taxonomique de la famille des Retroviridae.

L'alignement des séquences génomiques rétrovirales ainsi que la comparaison des séquences protéiques de la RT, l'élément protéique le plus conservé chez les rétrovirus, ont permis à l'ICTV (International Committee on Taxonomy of Viruses) de définir une classification des rétrovirus en 7 genres pouvant se répartir en deux sous-familles, que sont les *Orthoretrovirinae* (00.061.1) et les *Spumaretrovirinae* (00.061.2) (cf. figure 2).

Concernant la sous-famille des *Orthoretrovirinae*, six genres ont été déterminés :

<i>Alpharetrovirus</i>	}	rétrovirus simples	(00.061.1.03)
<i>Betaretrovirus</i>			(00.061.1.01)
<i>Gammaretrovirus</i>			(00.061.1.02)
<i>Epsilonretrovirus</i>			(00.061.1.08)
<i>Deltaretrovirus</i>	}	rétrovirus complexes	(00.061.1.05)
<i>Lentivirus</i>			(00.061.1.06)

Pour ce qui est de la sous-famille des *Spumaretrovirinae*, il n'y a qu'un seul genre, celui des *Spumavirus* (00.060.0.07) qui ne comprend que des rétrovirus non-oncogéniques. Ces virus établissent des infections persistantes chez différentes espèces animales (primates et humains (Primate Foamy Virus de type 1 : 00.061.2.07.003.00.002), chats, hamsters...). Ils sont considérés comme non pathogènes. Les *Spumavirus* peuvent être également considérés comme des rétrovirus complexes puisque leur génome possède deux ORFs supplémentaires codant pour des protéines accessoires.

## **I.B. Le genre Lentivirus**

Les virus appartenant à ce genre causent des maladies généralisées, caractérisées par de longues périodes d'incubation (jusqu'à 11 ans pour le VIH-1) et par une infection persistante. Ils infectent différentes espèces animales dont les primates (VIH-1, VIH-2, SIV), les félins (FIV) ou encore les bovins (BIV). Les particules virales possèdent un diamètre de 80 à 100 nm et les protéines d'enveloppe sont glycosylées. La capside est de forme conique. Les lentivirus sont des rétrovirus complexes. Leur génome dont la taille est supérieure à 9 kb (9,2 kb dans le cas du VIH-1) code pour les polyprotéines rétrovirales classiques Gag, Pol et Env et pour différentes protéines auxiliaires et régulatrices (Tat, Rev, Nef, Vif, Vpr et Vpu chez le VIH-1) au niveau d'« Open Reading Frames » (ORFs) situés entre le gène pol et le gène env ainsi que dans la région 3' du gène env (cf. figure 11 A : organisation du génome du VIH-1).

Les lentivirus se différencient des autres rétrovirus par leur capacité à infecter des cellules sans qu'elles entrent nécessairement en phase de mitose. Le passage de la paroi nucléaire est réalisé alors que le génome viral interagit avec différents partenaires au sein d'un complexe supramoléculaire de préintégration (CPI) nucléoprotéique. Cette entité passe dans le noyau par un mécanisme qui n'est pas encore très clair. Plusieurs acteurs moléculaires du CPI ont été identifiés comme ayant un rôle dans son adressage nucléaire.

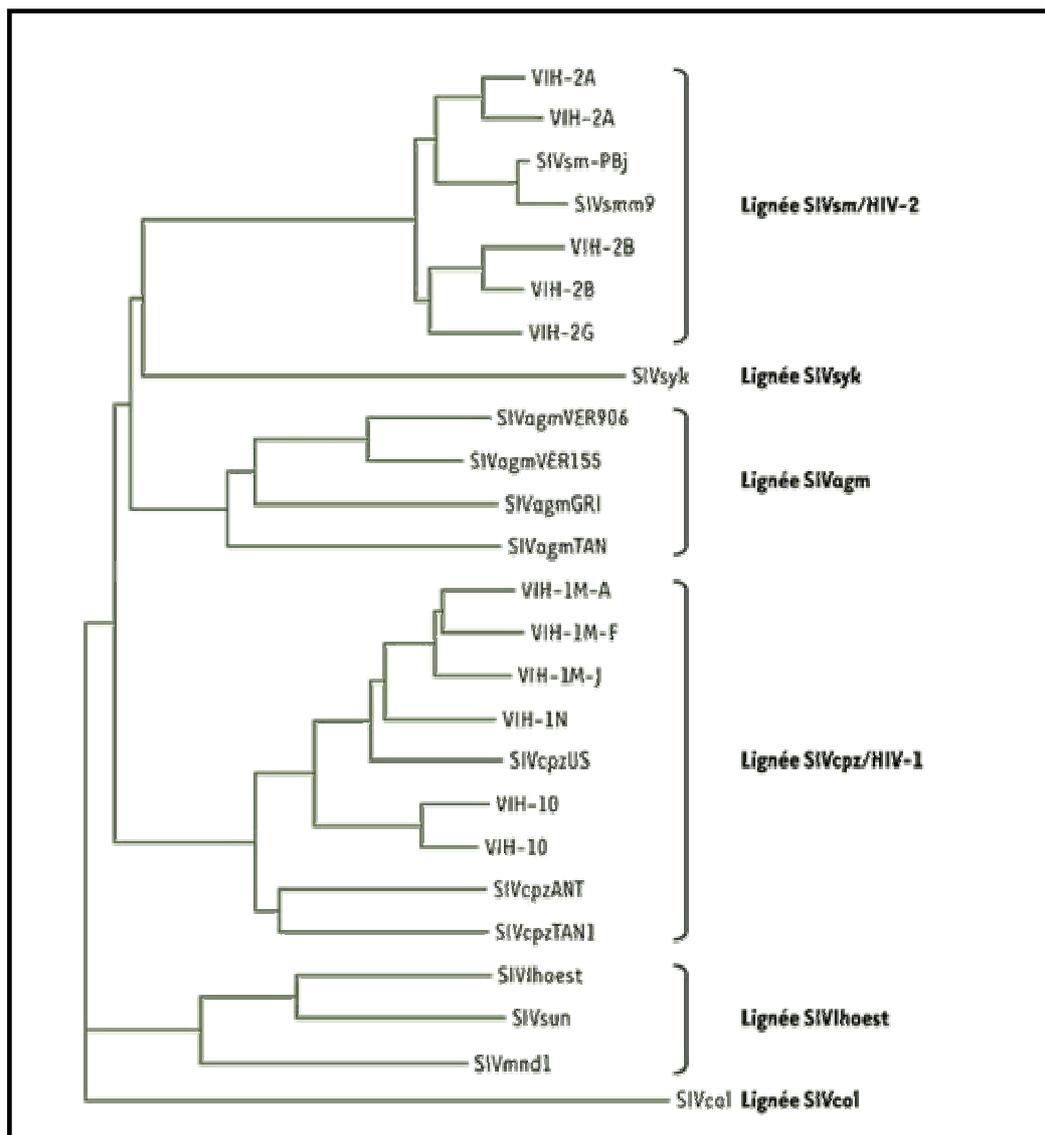
# Le VIH-1

## *1.B.1. Présentation, historique et origine du VIH-1*

Le Virus de l'Immunodéficience Humaine 1 (VIH-1) possède un tropisme dirigé contre les cellules du système immunitaire. Il infecte de manière prépondérante les lymphocytes T CD4+ (LT CD4+), les macrophages, les cellules dendritiques, ainsi que des cellules des systèmes lymphatique, hématopoïétique et nerveux. La destruction des LT CD4+, qui constituent la clé de voûte du système immunitaire, va conduire à une chute de l'efficacité de la réponse immunitaire face aux agressions extérieures. En dessous d'un certain taux de LT CD4+, différentes infections opportunistes vont pouvoir se développer. Au stade inférieur à 300 LT CD4+ par  $\mu\text{L}$  de sang, la personne infectée développe le Syndrome d'Immunodéficience Acquise (SIDA).

Le virus fut mis en évidence aux Etats-Unis en 1981 à partir de l'observation d'une recrudescence d'épidémies d'infections opportunistes, habituellement observées chez des individus fortement immunodéprimés. La population touchée se composait principalement d'homosexuels, de toxicomanes et de personnes ayant subi une transfusion sanguine. Les patients présentaient une diminution très importante du nombre de LT CD4+. Cette affection fut associée au SIDA. En 1983, l'équipe du Professeur Montagnier (Institut Pasteur, Paris, Prix Nobel de Médecine 2008) parvient à isoler un rétrovirus à partir de ganglions d'un patient présentant une lymphadénopathie (Barre-Sinoussi *et al.*, 1983). Il est dénommé LAV (lymphoadenopathy-associated virus) et ses caractéristiques se révèlent être proches de celles du HTLV (Human T cell Lymphotropic Virus) découvert en 1980 par le Professeur Gallo (Etats-Unis) (Poiesz *et al.*, 1980). En 1986, une commission scientifique propose une dénomination commune : le virus responsable du SIDA est appelé VIH-1. L'équipe du Pr. Montagnier identifie en 1986 un deuxième virus, le VIH de type 2, chez des patients originaires d'Afrique de l'Ouest. Ces deux virus présentent des symptômes cliniques ainsi que des propriétés structurales, génétiques et biologiques similaires (Clavel *et al.*, 1986).

La pandémie et la mortalité dues aux deux VIH sont telles, que beaucoup de chercheurs se sont interrogés quand à son origine. Différentes études montrent, notamment grâce à la phylogénie que les deux VIH possèdent une origine animale ((Sharp *et al.*, 1999) ; (Courgnaud *et al.*, 2004)). En effet, plus d'une trentaine d'espèces de singes africains sont



**Figure 3 : Arbre phylogénétique des VIH-1 et -2**

Arbre phylogénétique créé à partir des séquences pol des différents VIS dont deux lignées comprennent le VIH-1 et le VIH-2. D'après (Cournaud et al., 2004)

connues pour être infectées par le VIS (Virus de l'Immunodéficience Simien) mais de manière asymptomatique. La construction de l'arbre phylogénétique des lignées de VIS à partir des séquences protéiques Pol des différents lentivirus de primates permet de dégager six lignées principales (cf. figure 3) dont :

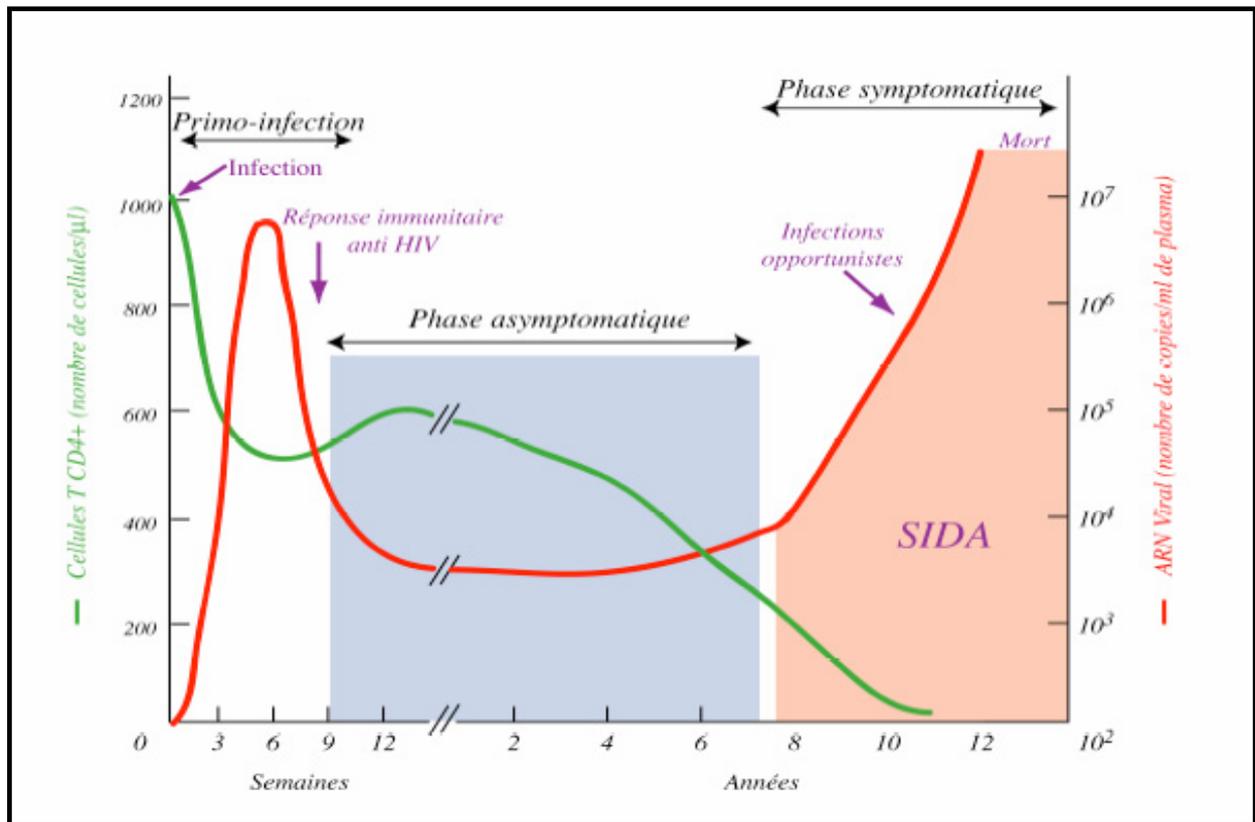
- une lignée VIScpz (retrouvée chez le chimpanzé *Pan troglodytes troglodytes*) fortement associée au VIH-1
- une lignée VISsm (retrouvée chez le mangabey enfumé *Cercocebus torquatus*) qui est associée au VIH-2.

Les deux souches de VIH se différencient donc de part leur lien de parenté avec le VIS, mais aussi par leur origine géographique ainsi que par l'organisation de leur génome. Le VIH-2 se retrouve principalement dans l'ouest africain, alors que le VIH-1 présente une répartition mondiale. Il y a trois sous-groupes de VIH-1 selon les critères phylogénétiques : les groupes M (le plus courant), N et O eux même subdivisés en sous types ou clades. Afin d'expliquer cette origine animale, deux hypothèses d'évolution sont envisageables. D'une part, il existerait une co-évolution entre le virus et l'hôte sans qu'il y ait de transfert interspèce. Ceci est notamment illustré par les quatre VISagm différents isolés chez chacune des quatre espèces principales de singe vert différenciées par des critères morphologiques, génétiques et géographiques. D'autre part, la parenté virale peut suivre la proximité géographique et dans ce cas, l'hypothèse d'une transmission interspèce semble plus logique. Ce serait le cas des deux types de VIH qui sont très proches des virus simiens présents dans les mêmes zones géographiques.

### ***1.B.2. Pathogénicité du VIH-1***

Chez l'homme, la cinétique d'infection du VIH-1 peut-être divisée en trois phases successives : la primo-infection, la phase de latence clinique et la phase SIDA précédant la mort (cf : figure 4).

La primo-infection est cliniquement associée à des symptômes proches de ceux observés pour une grippe ou une mononucléose. Ils apparaissent approximativement 3 à 6 semaines après l'infection. Le nombre de LT CD4+ circulants dans le sang baisse de manière significative dans les 4 à 8 semaines. Ce niveau peut remonter lors de l'entrée en phase de latence clinique mais rarement au niveau observé avant l'infection. Les symptômes associés à



**Figure 4 : Cinétique de l'infection au VIH**

Les trois phases d'évolution d'une infection au VIH reliées au nombre de LT CD4+ circulants et à la charge virale du plasma.

la primo-infection sont accompagnés d'une répllication virale intense détectable dans le sang à partir de la troisième semaine post-infection. Durant cette période, les protéines ainsi que les particules virales peuvent être détectées dans le plasma et dans les liquides cérébrospinaux. Le nombre de virus plasmatiques peut atteindre  $10^6$  à  $10^7$  particules par millilitre (ml). Cependant, les particules infectieuses se rencontrent à une fréquence de 1 sur 3000 à 1 sur 100 000 (Piatak, Jr. *et al.*, 1993). La réponse immunitaire dirigée contre le VIH commence à être efficace à partir de la troisième semaine (Graziosi *et al.*, 1993) (cf. figure 4).

La chute des antigènes viraux observable est associée à une diminution des symptômes cliniques et à une stabilisation du nombre de LT CD4+. Cette période, nommée période de latence clinique, est caractérisée par une absence de symptômes. Durant cette phase, un déclin progressif des LT CD4+ est observé sauf chez les individus dits non-progresseurs. Bien que cette phase asymptomatique soit appelée phase de latence clinique, plusieurs indices démontrent qu'il y a une répllication virale rapide et continue durant cette période. En effet, les ARNs viraux peuvent être détectés par hybridation *in situ* à partir de ganglions lymphatiques et par PCR (Polymerase Chain Reaction) à partir de plasma de patient à tous les stades de la maladie ((Pantaleo *et al.*, 1993); (Piatak, Jr. *et al.*, 1993)). Mais cette répllication est, pour un temps, contrôlée par le système immunitaire.

La phase de latence clinique permet la dissémination du virus dans tout l'organisme hôte. Macrophages, monocytes, cellules dendritiques et cellules nerveuses seront infectées par le VIH durant cette période. Les macrophages sont particulièrement impliqués dans ce processus puisqu'ils favorisent le passage du virus à travers les barrières hématoencéphalique et alvéolocapillaire. Le VIH n'étant pas cytopathique et se répliquant faiblement à l'intérieur des macrophages (par rapport aux LT CD4+), et ces cellules ayant une demi-vie plus longue (14 jours au lieu de 1,5 jours pour les LT CD4+), ils pourraient servir de cellules réservoir ((Cavert and Haase, 1998) ; (Ho, 1998)). Certains LT CD4+ quiescents infectés par des particules virales défectives ou contenant un ADN viral latent intégré ou non-intégré dans leur génome constituent également des réservoirs potentiels pour le VIH-1 (Finzi *et al.*, 1997).

Après plusieurs mois ou années, le niveau des LT CD4+ passe du nombre normal compris entre 600 et 1200 cellules par millilitre (mL) à moins de 500 cellules / mL (cf. figure 4). Les symptômes cliniques associés au SIDA peuvent alors être observés. La troisième phase de la maladie appelée SIDA est définie par un nombre de LT CD4+ inférieur à 300

**Figure 5 : Cycle de réplication du VIH-1**

Les grandes étapes de la réplication du VIH-1 et celles potentiellement inhibées par des drogues commercialisées (a, b, c et e) ou à l'étude (d et e).

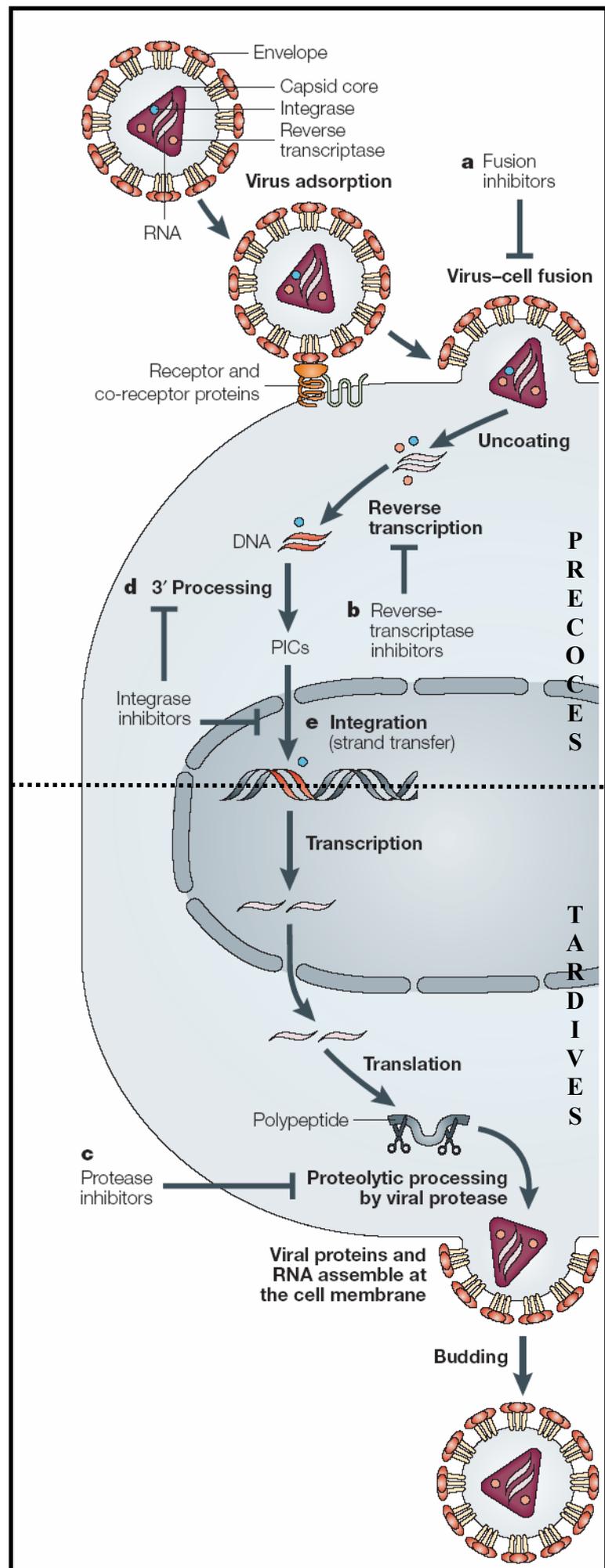
**a :** Inhibiteurs de fusion bloquant l'entrée du virus dans la cellule.

**b :** Inhibiteurs de la transcriptase inverse bloquant la synthèse de l'ADN viral.

**c :** Inhibiteurs de la protéase bloquant la maturation en inhibant la protéolyse des polyprotéines virales catalysée par la protéase.

**d et e :** Inhibiteurs de l'IN. 2 types d'inhibiteurs, ceux du 3'processing (maturation de l'ADN viral, **d**) et ceux du transfert de brin (intégration de l'ADN viral dans l'ADN hôte, **e**)

*D'après (Pommier et al., 2005)*



cellules / mL. Durant cette phase, l'individu infecté devient sensible à des infections dites opportunistes. Ces dernières touchent essentiellement les poumons, le cerveau, les yeux et le tube digestif. Les plus fréquentes sont la pneumocystose (infection pulmonaire due à *Pneumocystis jiroveci*), les candidoses oesophagiennes (infections de l'oesophage dues à des champignons appartenant au genre *Candida*), la tuberculose (due à *Mycobacterium tuberculosis*), le sarcome de Kaposi (du au Virus HHV-8) et les infections à cytomégalovirus dit CMV ainsi que différentes autres infections. Le virus circulant devient de nouveau apparent et l'augmentation de ces titres est corrélée avec le stade avancé de la maladie (Ho *et al.*, 1989). La corrélation entre la sévérité des symptômes et le nombre total de LT CD4+ est bien établie et représente toujours le critère par excellence de pronostic.

### ***I.B.3. Cycle répliatif du VIH-1***

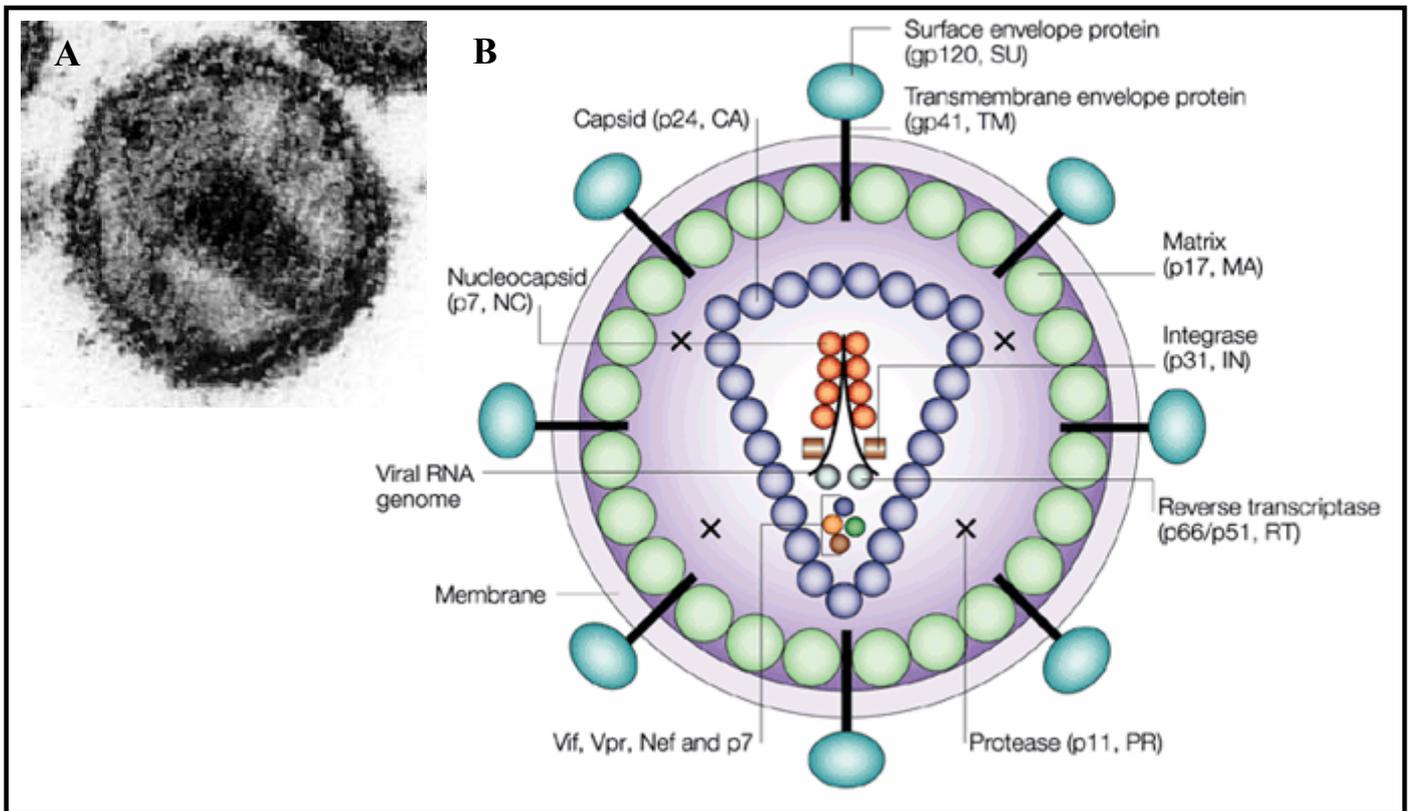
Les grandes étapes du cycle répliatif du VIH sont décrites sur la figure 5 ci-contre. Les phases précoces du cycle de répliation débutent par la reconnaissance de la cellule cible par le virion et vont jusqu'à l'intégration du génome viral au sein du génome de la cellule hôte. Les étapes tardives commencent avec l'expression régulée du génome proviral intégré et tout les processus mis en route pour arriver à la libération dans le milieu extérieur et la maturation des virions néoformés.

#### **I.B.3.a. Structure et composition du virion**

La morphologie et la structure du VIH-1 ont été résolues par microscopie électronique (Gelderblom *et al.*, 1987) (cf. figure 6A). Les particules virales mûres sont grossièrement sphériques et possèdent un diamètre d'environ 100 nm. La capsid est incluse dans une membrane à bicouche phospholipidique à la surface de laquelle des spicules composés de deux glycoprotéines (cf. figure 6B) sont observés :

- SU (Surface Unity) gp120 qui se fixe au récepteur cellulaire.
- TM (TransMembrane) gp41, liée à gp120, qui est responsable de la fusion de l'enveloppe virale avec la membrane cellulaire.

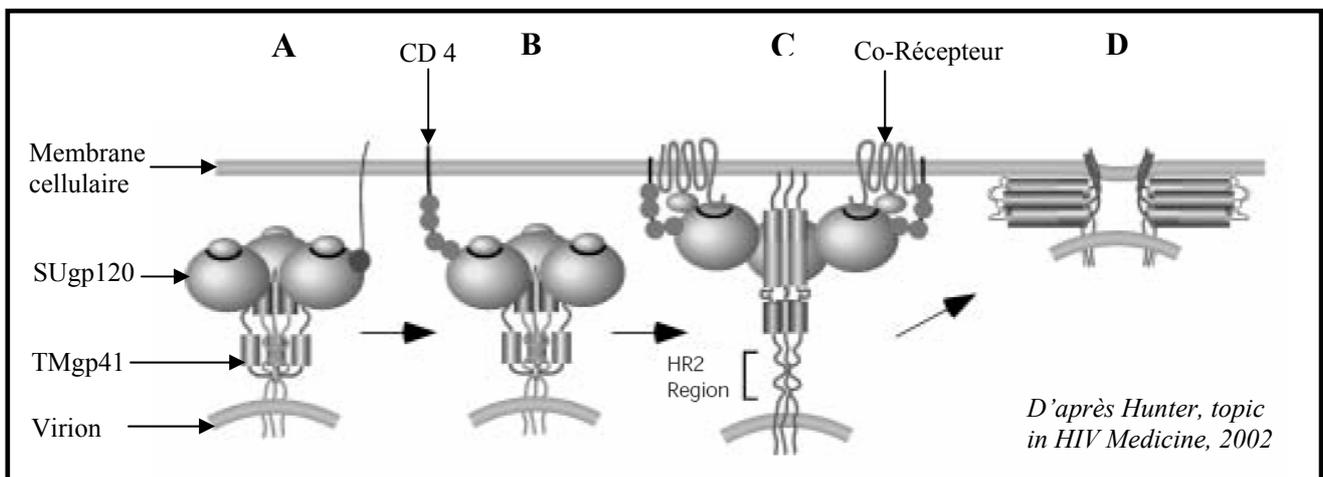
Ces deux protéines (membranaire pour SUgp120 et transmembranaire pour TMgp41) adoptent un regroupement en trimères.



**Figure 6 : Structure du virion mature du VIH-1**

**A** : Virion mature du VIH-1 au microscope électronique, la zone dense centrale est la capsid qui contient le génome viral.

**B** : Organisation et composition d'un virion mature. *D'après (Freed, 1998)*



**Figure 7 : Processus d'entrée du VIH-1 dans la cellule**

**A** : Attachement du virion à la membrane cellulaire par des liaisons non spécifiques

**B** : Liaison au récepteur CD4, initiation des changements conformationnels

**C** : Liaison au corécepteur, achèvements des changements conformationnels avec ancrage de la membrane du virion à la membrane cellulaire par TMgp41.

**D** : 2<sup>ème</sup> changement conformationnel de TMgp41 avec un repliement en « hairpin » qui permet le rapprochement des deux membranes et leur fusion.

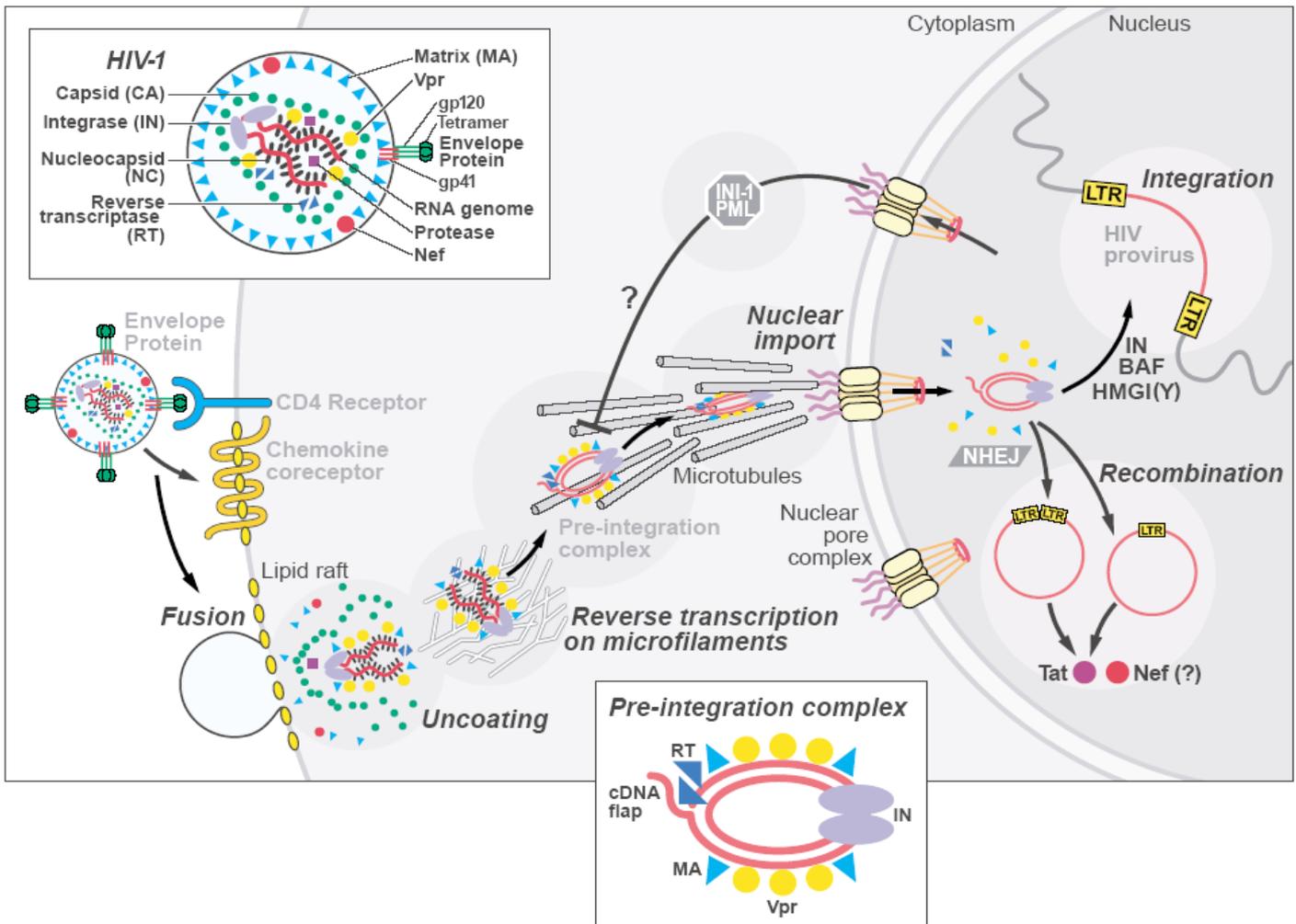
La membrane interne est tapissée par environ 2000 copies de la protéine de la matrice (MA). Au centre, la capside de forme conique est formée par l'assemblage d'environ 2000 copies de la protéine CA (p24). Les enzymes présentes dans le virion sont la transcriptase inverse (RT), l'intégrase (IN) et la protéase (PR). Une cinquantaine de molécules de chaque enzyme est présente dans chaque virion. Les particules virales contiennent également trois des six protéines accessoires virales que sont Nef (Negative factor), Vif (Virion Infectivity Factor) et Vpr (Viral Protein Regulatory). Le génome est composé de 2 ARNs monocaténares et contient 9 gènes différents qui codent pour 15 protéines virales. Les deux ARNs sont étroitement associés, en un complexe ribonucléoprotéique stabilisé par les protéines de la nucléocapside (NCp7) (environ 2000 copies).

Les trois autres protéines accessoires que sont Rev (Regulator of Viral Expression), Tat (Trans-Activator for viral Transcription) et Vpu (Viral Protein Unknown) n'apparaissent pas dans les virions matures.

### **I.B.3.b. Les étapes précoces du cycle de réplication**

Le processus d'entrée dans la cellule cible requiert des interactions spécifiques entre l'enveloppe virale et des récepteurs de surface cellulaire. Les deux protéines d'enveloppe virale (SUgp120 et TMgp41) sont associées de manière non covalente sous forme de trimères qui forment l'unité fonctionnelle active (cf. figure 7A). Les trimères de SUgp120 interagissent avec le récepteur de surface CD4 de la cellule cible (cf. figure 7B). Cela induit un changement conformationnel des protéines d'enveloppe qui permet l'interaction avec des récepteurs secondaires spécifiques (Kwong *et al.*, 1998) (cf. figure 7C). A ce jour, deux d'entre eux jouent un rôle bien établi *in vivo* : CCR5 situé à la surface des macrophages et CXCR4 à la surface des lymphocytes T. CCR5 semble être utilisé invariablement lors de chaque nouvelle infection. Une délétion de 8 résidus au niveau d'un récepteur mutant CCR5 a été mise en évidence et les individus homozygotes pour cette délétion (1 à 2% de la population de type caucasien) sont totalement résistants au VIH. Cette observation a mis en lumière l'importance de ce corécepteur et a permis l'ouverture d'une nouvelle voie de recherche d'inhibiteur de la réplication du VIH.

L'interaction entre SUgp120, CD4 et le ou les corécepteurs induit un changement conformationnel radical de la sous unité transmembranaire TMgp41. Le trimère



**Figure 8 : Etapes précoces du cycle de réplication du VIH**

Décapsidation progressive après la fusion. La transcription inverse est réalisée pendant le trafic intracellulaire du CPI. L'ADN viral double brin est ensuite mûri par l'IN, puis transloqué dans le noyau. Après le passage des pores nucléaires, le génome viral est intégré.

A noter la formation des cercles à un et deux LTR.

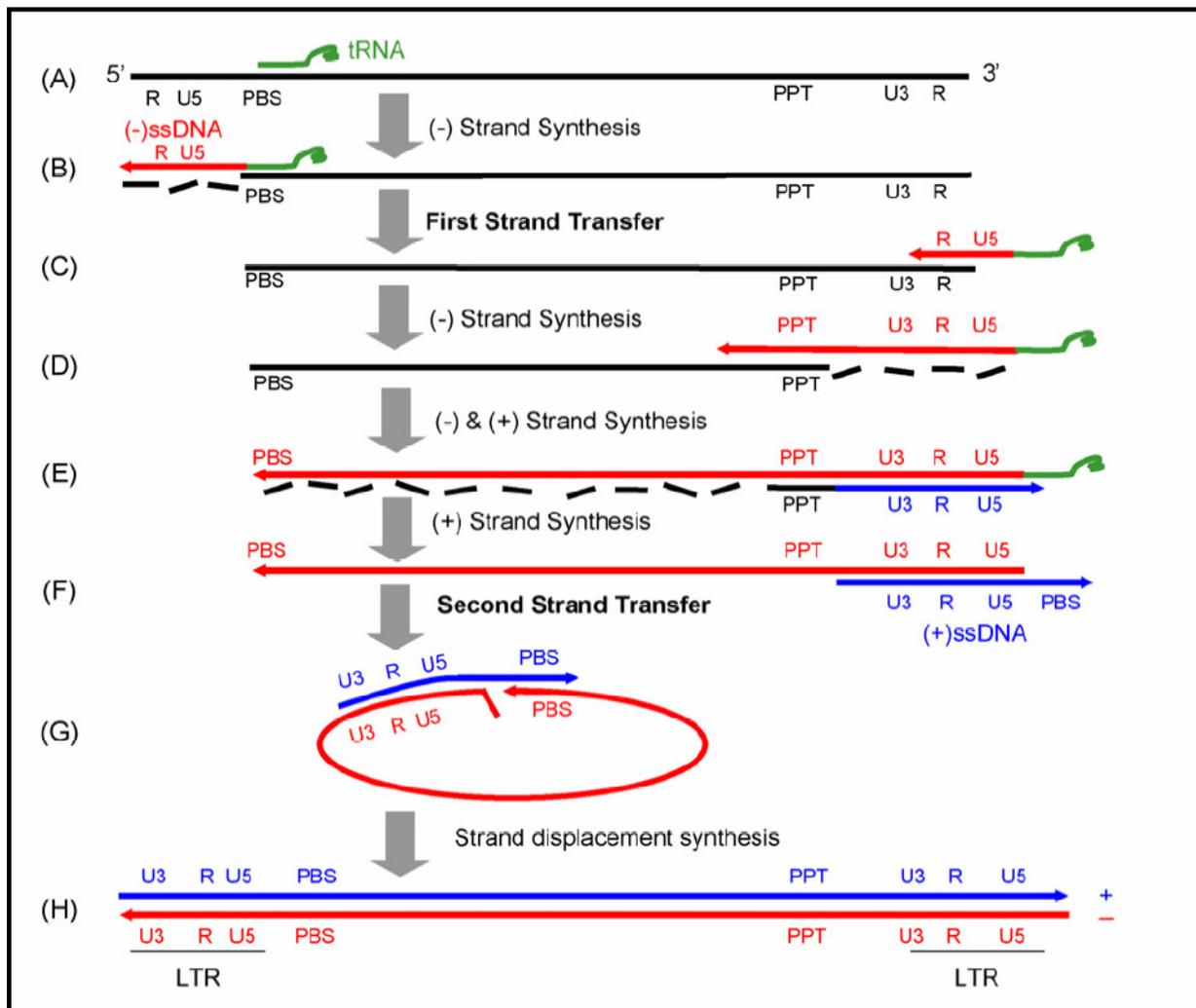
En *in set* : les composants viraux identifiés comme appartenant au CPI.

D'après (Greene and Peterlin, 2002)

de TMgp41, qui possède une structure tertiaire coil-coiled, projette ses trois peptides de fusion afin d'amarrer la bicouche lipidique de la cellule cible (cf. figure 7C). Un deuxième changement de structure de TMgp41 dit en épingle à cheveux (hairpin) permet le rapprochement des deux membranes et favorise ainsi la fusion et la libération de la capsid virale dans le cytoplasme cellulaire (Chan and Kim, 1998) (cf. figure 7D). Actuellement, il existe un inhibiteur de fusion (T-20) qui bloque le changement structural aboutissant à la formation de cette structure dite « hairpin ». Notons qu'un système d'entrée par endocytose a été également mis en évidence au niveau des cellules dendritiques (Geijtenbeek *et al.*, 2000). Le VIH-1, par l'intermédiaire de SUGp120, interagit avec des récepteurs membranaires spécifiques DC-SIGN. Les cellules dendritiques, lors de leur maturation, quittent les zones périphériques pour les ganglions lymphatiques. Arrivées à ce niveau, les cellules dendritiques exposent les virus fixés à DC-SIGN et permet donc à la manière d'un cheval de Troie d'infecter les lymphocytes T présents.

Lorsque la membrane du virion fusionne avec celle de la cellule hôte, la capsid est libérée dans le cytoplasme. Le mécanisme de sa déstructuration reste encore à élucider. Le processus pourrait intervenir après une phosphorylation de MA par une MAPKinase et par l'action de différentes protéines cellulaires et virales comme la cyclophiline A, Vif et Nef. La capsid désintégrée, le complexe de transcription inverse est libéré (cf. figure 8). Ce complexe inclut la double copie d'ARN génomique viral, l'ARNt de la lysine (ARN<sub>t3</sub><sup>lys</sup>) qui sert d'initiateur de la transcription inverse ainsi que d'autres facteurs dont Vpr et des facteurs cellulaires. Ce complexe est pris en charge au niveau des microfilaments d'actine par l'intermédiaire de MA phosphorylée. Cette interaction est obligatoire pour synthétiser l'ADN viral. A noter également l'importance de Vif pour l'étape de transcription inverse. Elle va limiter l'effet mutagène induit par la déaminase APOBEC3G en favorisant sa séquestration cytoplasmique lors de l'assemblage viral (Henriet *et al.*, 2005).

La transcription inverse est réalisée par la RT qui possède toutes les activités catalytiques nécessaires à la conversion de l'ARN génomique viral en ADN bicaténaire. D'une part, l'activité ADN polymérase ADN- et ARN-dépendante permet de synthétiser un génome ADN double brin à partir d'un ARN mono brin. D'autre part, l'activité endonucléasique ARN (ARNase H) est responsable de la dégradation de la matrice ARN et des amorces ARNt durant la synthèse d'ADN viral. Ces deux activités sont situées sur des domaines distincts de la RT (Kogel *et al.*, 1995). L'activité ADN polymérase est quelque peu différente de celle que l'on observe au niveau des systèmes répliatifs bactérien et eucaryote



**Figure 9 : Transcription inverse réalisée par la RT**

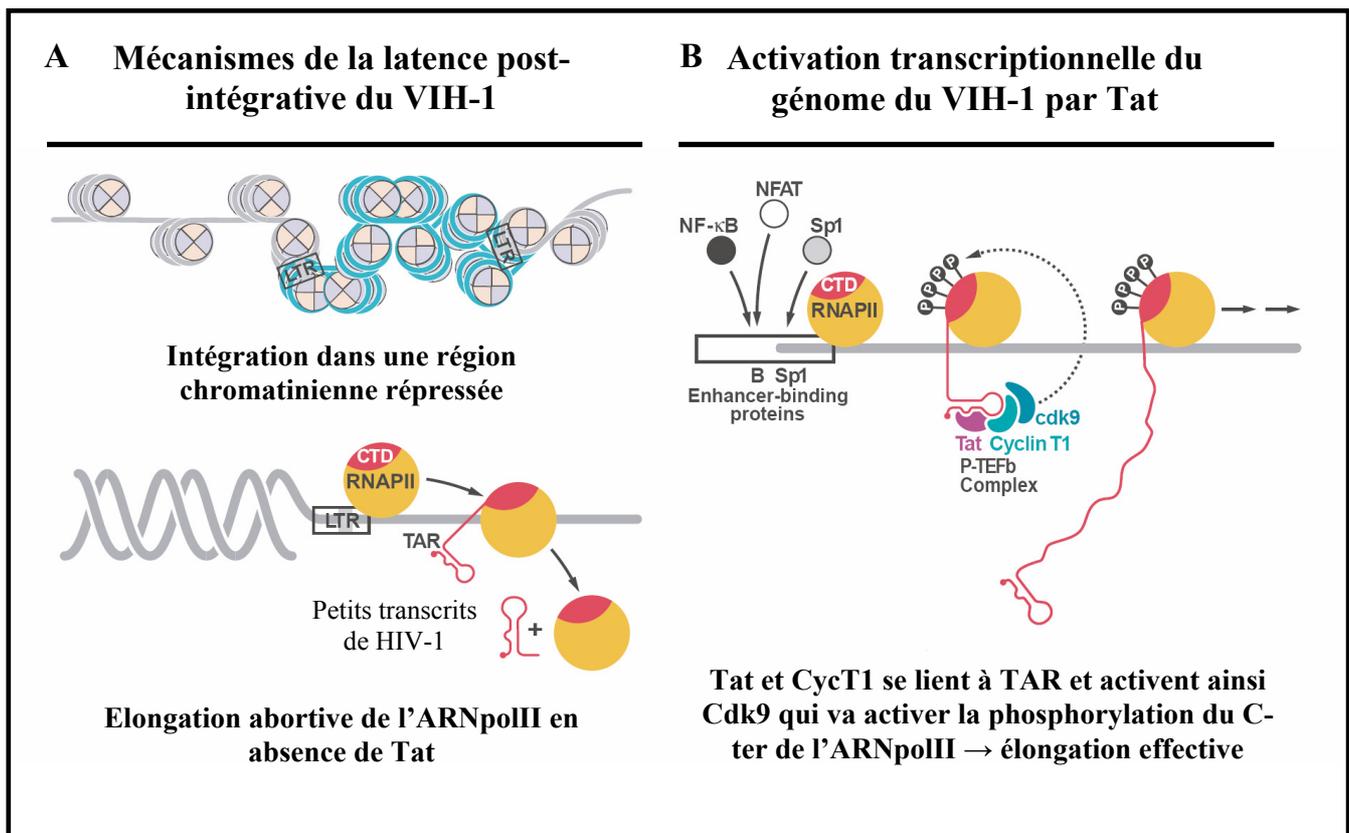
- A :** Initiation de la transcription inverse par un primer ARN (ARN<sub>13</sub><sup>lys</sup>), en vert, qui s'hybride partiellement avec la région PBS de l'ARN génomique, en noir.
- B :** Synthèse du début du brin d'ADN (-), en bleu, arrêt rapide. L'extrémité 5' de l'ARN viral est dégradée.
- C :** **1<sup>er</sup> saut de brin** : hybridation de la séquence R de l'ADN (-) avec celle de l'extrémité 3' de l'ARN viral.
- D :** Poursuite de la synthèse du brin ADN (-) avec dégradation de l'ARN viral parallèlement à la synthèse de l'ADN, à l'exception de la séquence PPT.
- E :** L'ARN PPT non dégradé sert d'amorce à la synthèse du brin (+) d'ADN viral, en rouge. L'ADN (+) est synthétisé jusqu'à l'extrémité U5.
- F :** La séquence PBS du brin (+) est synthétisée à partir de l'ARNt, puis après sa dégradation, on a le **2<sup>ème</sup> saut de brin**.
- G :** Poursuite de la synthèse du brin d'ADN (+).
- H :** La synthèse s'achève au niveau des extrémités LTR afin d'avoir 2 séquences terminales identiques (U3-R-U5 complètes).

*D'après (Basu et al., 2008)*

puisque'il y a absence de relecture (3'-5' exonucléasique) ce qui induit l'apparition de nombreuses mutations au cours de la réplication virale. Pour le VIH-1, cette fréquence est très élevée (1 base sur 10,000) (Bebenek *et al.*, 1993) ce qui fait environ une mutation par cycle de réplication. Le processus de la transcription inverse est réalisé en plusieurs étapes dont deux sauts de brins intra ou inter moléculaires qui vont permettre d'aboutir à un ADN viral double brin possédant des extrémités franches et avec à ces extrémités (LTR) des groupes de séquences identiques dans le même ordre : U3-R-U5 (cf. figure 9 pour un schéma détaillé du mécanisme).

Ce complexe nucléoprotéique est également appelé complexe de préintégration (CPI). Le CPI se compose du double brin d'ADN viral, de l'IN, MA, Vpr, RT, mais aussi de facteurs cellulaires dont les HMGI(Y) (High mobility group DNA binding cellular protein (Miller *et al.*, 1997) et LEDGF/p75 (Lens Epithelium-derived Growth Factor). Il se déplace dans le cytoplasme en suivant les microtubules selon un mouvement orienté rapide (0.1 à 1  $\mu\text{m}$  / sec) ((McDonald *et al.*, 2002), (Arhel *et al.*, 2006)). Au niveau de la région périnucléaire, le CPI atteint les pores nucléaires en suivant le réseau de filaments d'actine avec une vitesse inférieure à 0.03  $\mu\text{m}$  / sec (Arhel *et al.*, 2006). Sa translocation réalisée, le mouvement intranucléaire observé est alors typiquement diffusif et sans orientation préférentielle. Les Adénovirus et le virus de l'Herpes simplex utilisent également les microtubules pour gagner le noyau par l'intermédiaire de moteur moléculaire cellulaire comme la dynéine (cf. figure 8). Parallèlement à ce déplacement, l'IN va catalyser la maturation de l'ADN viral. Elle va cliver les deux extrémités 3' de l'ADN viral au niveau de la séquence consensus CAGT, entre le A et le G, libérant le dinucléotide GT. L'ADN viral ainsi mûré présentera alors des extrémités cohésives nécessaires à l'intégration future au sein du génome cellulaire hôte. Cette étape est appelée le 3'processing et elle sera décrite plus précisément dans la partie suivante.

L'ADN viral ainsi mûré pourra alors être intégré après son passage dans le noyau. Notons que l'ADN viral linéaire peut également se circulariser. Dans les cellules infectées, deux formes d'ADN viral circularisé coexistent. Ils sont dit cercles 1-LTR ou cercles 2-LTR selon qu'ils possèdent une (U3-R-U5) ou deux séquences LTR (U3-R-U5/U3-R-U5) juxtaposées (cf. figure 8). La circularisation de l'ADNc viral est réalisée selon un mécanisme qui n'est pas encore totalement élucidé. La présence des formes circulaires à 2-LTR de l'ADN viral peut, par exemple, être augmenté lorsque l'IN est catalytiquement inactive ou lorsque l'infection virale est réalisée en présence d'un inhibiteur de transfert de brin, qui bloque le



**Figure 10 : Régulation de la transcription post-intégrative du VIH**

**A** : Présentation des mécanismes favorisant la latence transcriptionnelle du provirus.  
**B** : Présentation de l'activation transcriptionnelle du provirus par la protéine accessoire Tat.

*D'après (Greene et al., 2002)*

cycle de réplication viral au niveau de l'étape de l'intégration. Le rôle, au niveau du cycle de réplication de ces formes circulaires de l'ADN viral, n'a pas encore été établi de manière claire (Brussel and Sonigo, 2004).

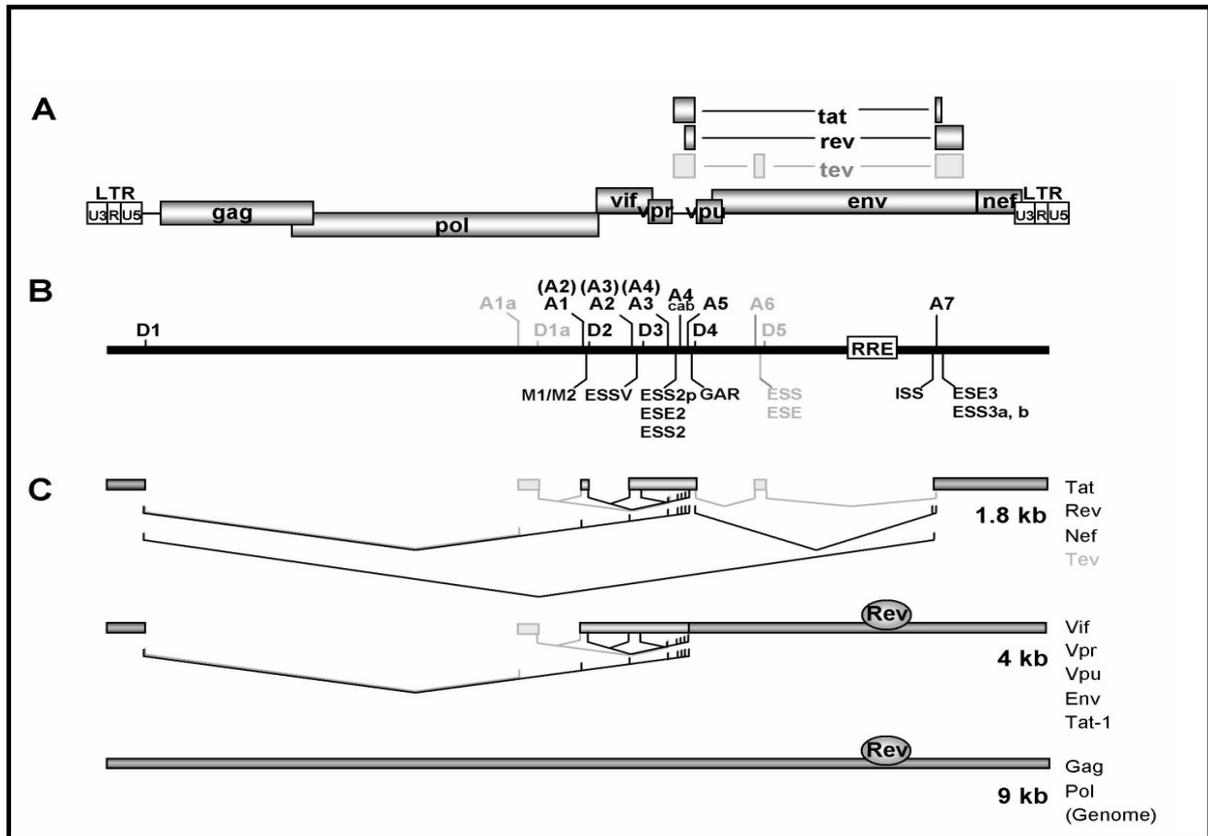
A l'inverse des autres rétrovirus, le VIH et les lentivirus peuvent infecter les cellules qui ne sont pas en phase de division. Cette étape oblige une traversée de la paroi nucléaire. Le complexe de préintégration présente un diamètre d'environ 50nm qui est le double du diamètre intérieur des pores nucléaires. De ce fait, sa traversée ne peut se faire par simple diffusion, mais nécessite un mécanisme actif dont les différentes phases et intervenants restent encore flous. Parmi les différents partenaires impliqués dans le CPI, plusieurs séquences nécessaires à la translocation nucléaire ont été mises en évidence que se soit au sein de l'IN (Gallay *et al.*, 1997), MA (Bukrinsky *et al.*, 1993) et Vpr (Heinzinger *et al.*, 1994). De plus, il est possible que l'ADN proviral lui-même intervienne car il présente dans sa structure un 'DNA Flap' de 99 nucléotides qui pourrait constituer en lui-même un signal de localisation nucléaire (Nermut and Fassati, 2003) (cf. Figure 8, *in set*). Le nombre et la diversité des séquences de localisation nucléaires (NLS) contenus au sein des différents éléments du CPI pourraient agir soit dans un système de coopération, soit jouer chacun un rôle plus important selon le type de cellules infectées (revue: (Bukrinsky, 2004)).

### **I.B.3.c. L'intégration**

La barrière nucléaire passée, le CPI peut intégrer l'ADN viral bicaténaire au sein du génome de la cellule hôte et ainsi devenir un provirus fonctionnel. Cette réaction est médiée par l'intégrase qui lie les extrémités de l'ADN viral au génome cellulaire (Miller *et al.*, 1997). Les protéines HMGI (Y) de la cellule hôte et BAF (barrière à l'autointégration) sont nécessaires pour avoir une intégration efficace, bien que leurs fonctions précises demeurent inconnues (Chen and Engelman, 1998). L'IN catalyse alors la deuxième réaction qui est l'intégration du génome viral et qui établit donc le provirus dans le génome cellulaire. Le mécanisme de l'intégration concertée sera décrit plus précisément dans la partie suivante (cf. figure 8).

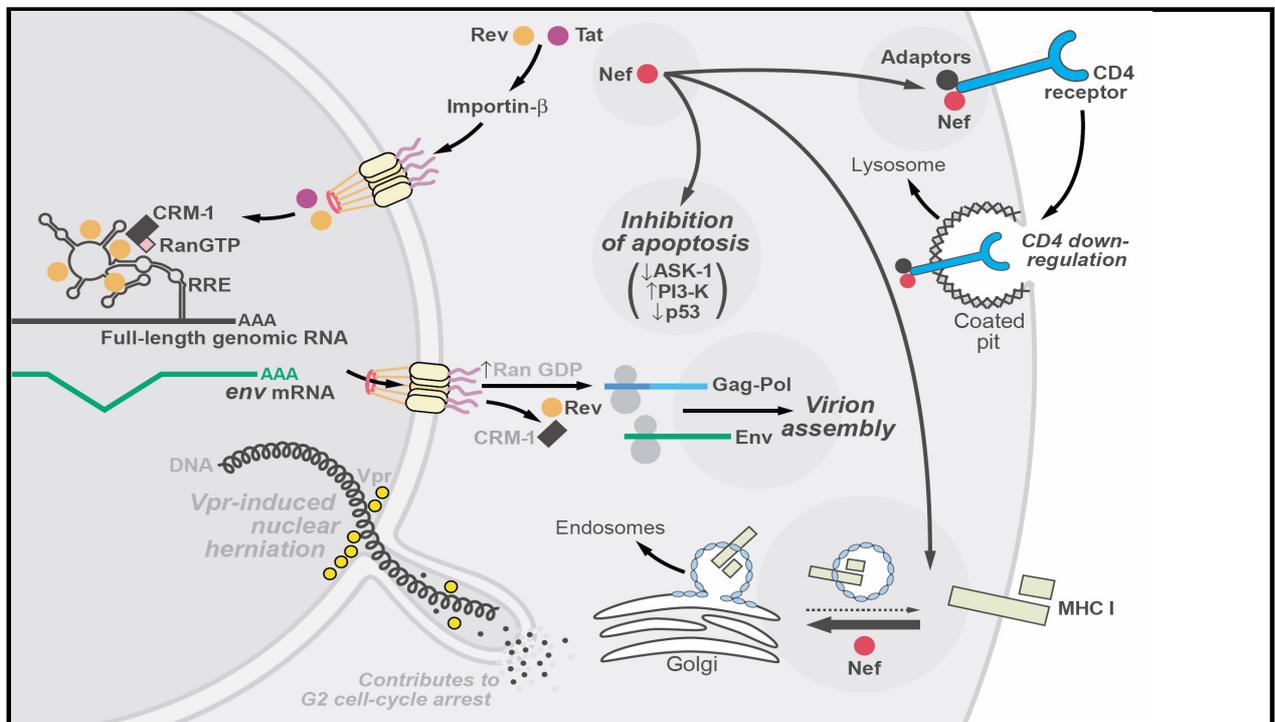
### **I.B.3.d. Les étapes tardives du cycle de réplication**

La latence transcriptionnelle du VIH explique l'impossibilité d'éradiquer définitivement le virus des patients infectés par des traitements basés sur des inhibiteurs



**Figure 11 : Epissage alternatif de l'ARNm du VIH-1**

A: Organisation du génome. B: Les sites d'épissages en 5' sont noté D1-5 et en 3' A1-6. RRE (Rev Responsive Element) ESS (Exon Splicing Silencer) et ESE (Exon Splicing Enhancer) modulent l'épissage. C: ARNm épissés, en présence ou non de Rev et les protéines traduites résultantes. D'après (Kammler et al., 2006)



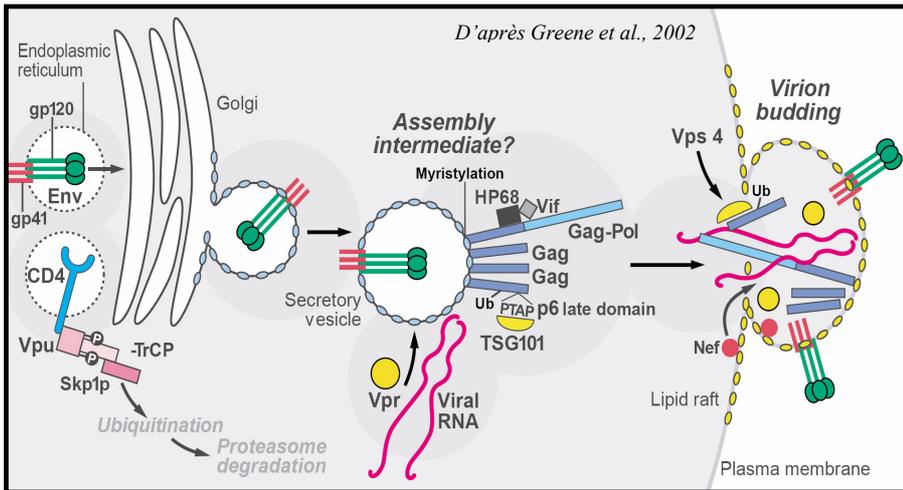
**Figure 12 : Influence de Nef sur l'environnement cellulaire**

Nef inhibe les processus apoptotiques et favorise la libération des virions. Vpr favorise l'arrêt du cycle en G2/M, et bloque les processus apoptotiques. D'après (Greene et al., 2002)

antirétroviraux. Actuellement, deux mécanismes potentiels de répression de la transcription virale sont étudiés. Le premier concerne l'environnement génomique au sein duquel le provirus est intégré : si le génome viral est dans une zone hétérochromatinienne qui est réprimée, sa transcription est également réprimée (Jordan *et al.*, 2001). La deuxième cause de latence peut être l'absence d'un enhanceur de transcription du VIH ou le déficit en Tat (cf. figure 10 A). Au niveau de la région 5'-LTR du provirus, différents éléments promoteurs sont présents : l'élément initiateur, la TATA box et trois sites Sp1 qui permettent le recrutement et le positionnement correct de l'ARN polymérase II (ARN pol II). Le génome du VIH possède également des enhanceurs transcriptionnels qui lient des facteurs nucléaires comme NFκB ou NFAT. La transcription commence, mais en l'absence de Tat, elle échoue rapidement (cf. figure 10 A) dès la fin de la synthèse d'un petit ARN non-polyadénylé formant une structure en tige-boucle stable. Avec la Cycline T1 (CycT1), Tat lie à petit ARN et recrute la kinase cycline-dépendante cellulaire 9 (Cdk9) qui va phosphoryler le domaine C-terminal de l'ARN pol II, permettant ainsi la poursuite de l'élongation de la transcription (cf. figure 10 B).

La réplication rétrovirale exige une exportation nucléaire de l'ARNm provenant d'un épissage alternatif de l'ARN viral génomique. L'ARNm non épissé de 9 kb code pour Gag ou Gag-Pol, quatre ARNm d'environ 4 kb sont épissés séparément. Ils codent pour Vif, Vpr, Vpu, et Env. Enfin, trois ARNm sont produits par multi-épissage. Ils aboutissent à des ARNm de 2 kb environ et codent pour Tat, Rev et Nef (cf. figure 11 C). Rev est la protéine virale qui contrôle les différents épissages de l'ARNm. En effet, en l'absence de Rev, les ARNm viraux transcrits seront entièrement épissés et aboutiront des ARNm de 2kb. En présence de Rev, il existe au niveau de l'ARNm viral en 3' une structure secondaire qui permet l'interaction de Rev avec une séquence ARM (Arginin Rich Motif) (cf. figure 11 B) (Cullen, 1998) ; (Malim *et al.*, 1989)). Les ARNm produits dans ce contexte seront ceux de 4 et 9 kb.

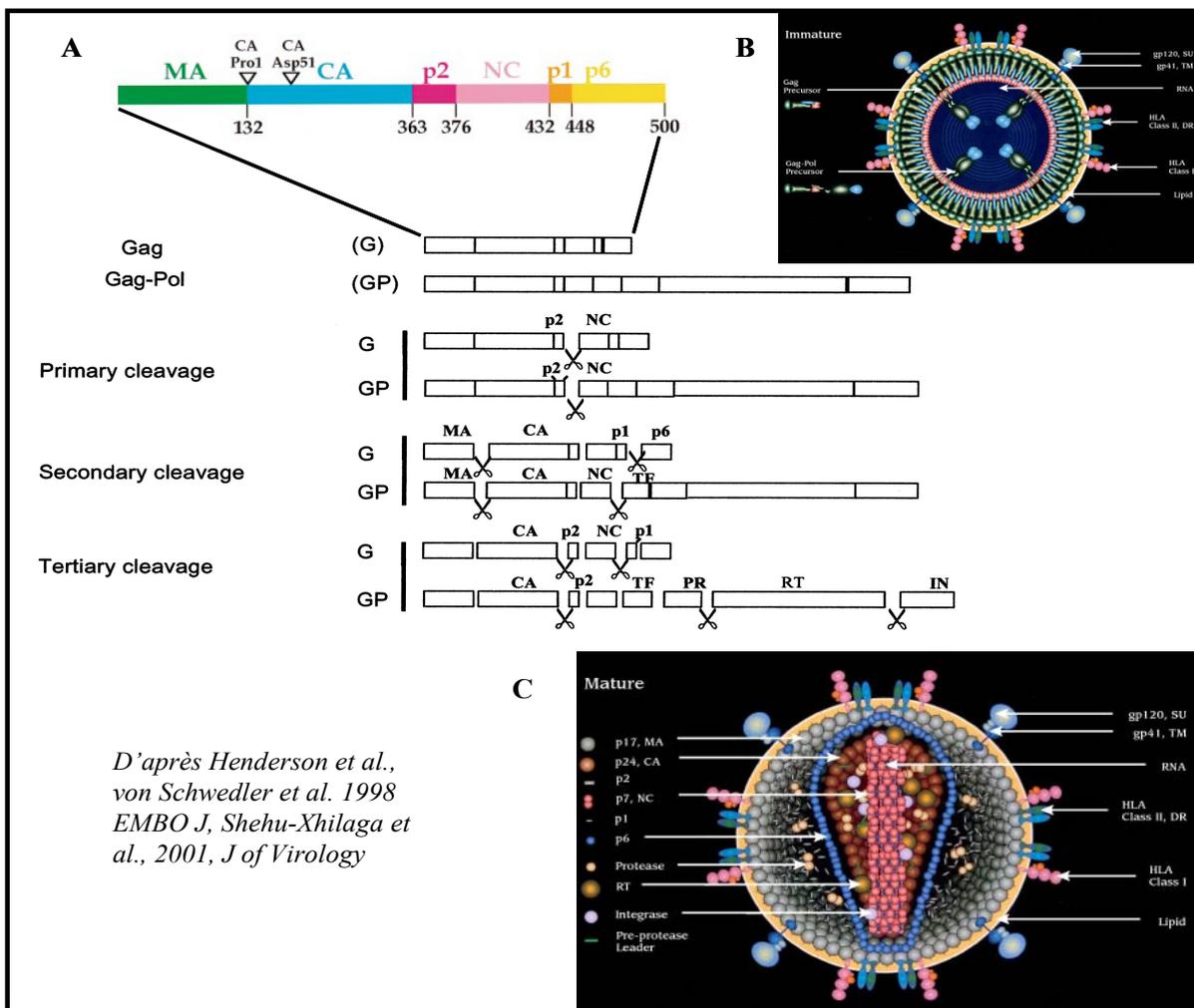
Alors que Rev et Tat interagissent avec l'ARN viral, Nef agit sur l'environnement cellulaire afin de favoriser l'assemblage et la libération de virions infectieux. Nef interfère avec plusieurs cascades de signalisation dont celle de l'activation des lymphocytes T et celle responsable de l'expression à la membrane des récepteurs CD4 et du CMH I. De plus, Nef inhibe les processus apoptotiques cellulaires en bloquant la cascade de signalisation au niveau de ASK-1 (Apoptosis Signal regulating Kinase-1). Nef interagit également avec le facteur antitumoral p53 qui peut être un initiateur apoptotique. Ainsi, Nef prolonge la vie des cellules infectées et favorise la production de virions (cf. figure 12).



**Figure 13 : Modifications post-traductionnelles et adressage à la membrane**

Les myristoylations de Gag et Gag-Pol permettent l'interaction avec les microdomaines où se réalise l'assemblage et la libération des virions immatures. L'interaction avec TSG101 permet l'adressage à la membrane

plasmique.



**Figure 14 : Maturation du virion du VIH par la protéase**

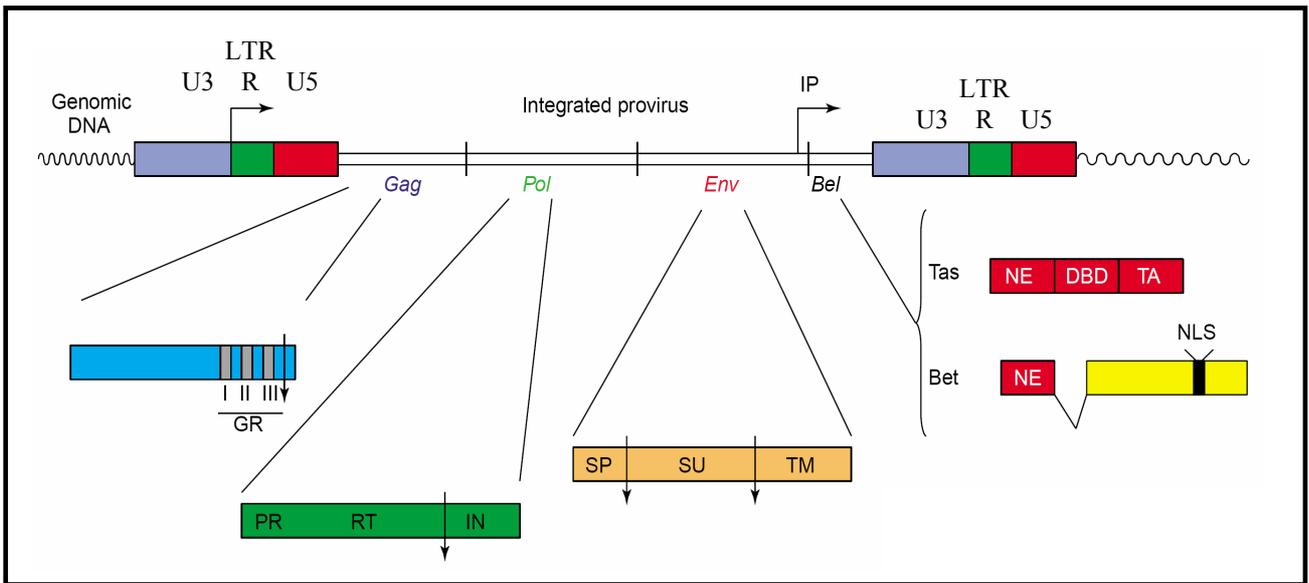
**A** : Clivages réalisés par PR sur Gag et Gag-Pol. 3 vagues successives pour aboutir au virion mature. 1<sup>ère</sup> vague : Séparation de NC ou NC-Pol, NC : les enzymes et l'ARN sont au centre du virion. 2<sup>ème</sup> vague : Séparation de CA pour former la capsid virale. 3<sup>ème</sup> vague : maturation de tous les composants de Gag et de Pol. *D'après (Shehu-Xhilaga et al., 2002)*

**B** : Virion immature. **C** : Virion mature.

Les nouvelles particules virales sont assemblées au niveau de la membrane plasmique. Chaque virion se compose d'approximativement 1500 molécules de Gag et de 100 polyprotéines de Gag-Pol, deux copies du génome viral d'ARN, et Vpr. Plusieurs protéines participent à l'assemblage, y compris les polyprotéines Gag et Gag-Pol, ainsi que Nef et Env. Une protéine cellulaire, HP68 (liant l'ATP), agit probablement comme une chaperonne moléculaire qui facilite le changement conformationnel de la polyprotéine Gag requis pour l'assemblage des capsides virales (cf. figure 13).

Les polyprotéines Gag et Gag-pol subissent des modifications post-traductionnelles permettant un adressage à la membrane dont des myristoylations qui leurs permettent d'interagir préférentiellement avec le cholestérol et les glycolipides qui composent préférentiellement les microdomaines membranaires. La maturation des polyprotéines Gag rend notamment apparent un domaine de type PTAP dans la séquence correspondant à la protéine virale p6 qui est reconnue par le TSG101 (expression du gène suppresseur de tumeur 101). Cette dernière est une protéine normalement associée aux protéines responsables de l'export des endosomes tardifs vers la membrane plasmique (cf. figure 13).

La maturation du virion est en majorité réalisée par clivage des polyprotéines Gag et Gag-Pol par la protéase du VIH-1 (PR) (cf. figure 14). Cette enzyme virale est une protéase à aspartate dont la forme active est un homodimère. L'interaction de dimères de Gag-Pol permet la reconstitution du site catalytiquement actif. PR va alors s'automaturer par clivage. La maturation du virion libéré se fait en plusieurs temps qui correspondent chacun à un type de clivage de PR. Les clivages de Gag vont aboutir aux protéines structurales MA, CA et NC, ceux de Gag-Pol : à MA, CA, NC ainsi qu'à PR, RT et IN. Les clivages de Gag ne se font pas au hasard, mais sont séquentiels (Wieggers *et al.*, 1998) (cf. figure 14 A). Dans un premier temps, PR libère NC, qui va former avec l'ARN viral le complexe nucléoprotéique central du virion. PR sépare ensuite MA de CA qui vont pour la première tapisser l'intérieur de la bicouche lipidique et pour la seconde former la capsidie conique qui contient l'ARN lui-même protégé par NC. Enfin, la maturation du virion s'achève lorsque PR catalyse le clivage des peptides situés aux extrémités de CA et NC (SP1 et SP2). Cette progression séquentielle est nécessaire au bon déroulement de la maturation des virions ((Pettit *et al.*, 2005), (Mirambeau *et al.*, 2007)). Dans le cas où la maturation ne s'effectue pas ou mal, les virions perdent tout pouvoir infectieux (figure 14 B et C).



**Figure 15 : Organisation génomique des spumavirus.**

Les protéolyses sont représentées par les flèches noires. IP est un promoteur interne. DBD est pour DNA Binding Domain (interaction avec l'ADN), NE pour séquence non essentielle et TA pour domaine TransActivateur, SP est le peptide signal de Env. A noter : pas de clivage entre la PR et la RT. Au niveau de Gag, présence des boîtes GR (I, II et III).  
 D'après (Delelis *et al.*, 2004)

## I.C. Le genre *Spumavirus*

Ce genre est particulièrement intéressant au niveau taxonomique car il présente toutes les caractéristiques des rétrovirus, mais il se distingue par d'autres qui rapprocheraient ces virus à des *pararétrovirus*, comme le virus de l'hépatite B (HBV) ou les rétrotransposons de levure. Ces virus furent isolés chez plusieurs mammifères comme le chat ou le cheval, mais ce sont surtout les primates dits non humains qui sont infectés par ce type de virus avec une prévalence de près de 70%. Quelques transmissions du virus primate / humains furent décrites, mais elles restent anecdotiques. En culture cellulaire, ces virus sont facilement identifiables car ils induisent la formation de syncytia, conduisant à la lyse cellulaire. Cependant *in vivo*, le virus semble être complètement dépourvu de pathogénicité tout en maintenant une infection persistante. Le prototype des *spumavirus* est le primate foamy virus de type 1 (PFV-1). Son étude a permis de mettre en évidence l'appartenance de ce genre aux rétrovirus dit complexes puisqu'en plus des gènes structuraux gag-pol-env, ils présentent deux ORFs additionnels qui, à la différence des *orthorétrovirus* complexes, se situent uniquement à l'extrémité 3' de leur génome (gène bel pour Between env and LTR), sous contrôle d'un promoteur interne transactivé par Bell (ou Tas) (cf. figure 15).

Seules les particularités propres aux *spumavirus*, qui les rapprochent notamment des *hepadnavirus* et des transposons de levure, seront exposées rapidement. Au niveau structural, leur génome diffère de celui des *orthorétrovirus* au niveau du gène gag. En effet, s'il code pour une nucléocapside, cette protéine virale ne contient pas de motif dit « boîte CH » caractéristique des repliements en doigt de zinc, qui permettent de former des interactions non spécifiques avec l'ARN génomique (Schliephake and Rethwilm, 1994). Ce dernier d'une longueur d'environ 12000 bases est le plus long de tous les rétrovirus. Lors d'une infection, contrairement aux autres rétrovirus, les virions des *spumavirus* ne sont retrouvés qu'en très faibles quantités dans le milieu extracellulaire. La majorité d'entre eux est localisée au niveau du réticulum endoplasmique des cellules infectées ((Goepfert *et al.*, 1995); (Hooks and Gibbs, Jr., 1975)). Il a également été montré que 20% des virions de PFV-1 contenaient de l'ADN (Yu *et al.*, 1996) contre 0,001% dans le cas du VIH-1 (Trono, 1992). Différentes données expérimentales permettent de penser que la transcription inverse est un mécanisme qui peut être précoce comme tardif dans le cycle de réplication des *spumavirus*, c'est-à-dire qu'elle peut avoir lieu avant ou après l'étape d'intégration et d'expression virale ((Moebes *et al.*, 1997);(Delelis *et al.*, 2003)). Par comparaison avec les *orthorétrovirus*, le génome

infectieux des *spumavirus* pourrait aussi bien être de l'ADN que de l'ARN ((Roy *et al.*, 2003); (Yu *et al.*, 1999)). Cette observation les rapprocherait des *hepadnavirus*, si ce n'était que l'intégration reste obligatoire dans le cycle de réplication des *spumavirus*.

Le déroulement de l'étape d'intégration diffère quelque peu de l'intégration observée par exemple pour le VIH-1. En effet, l'intégration comprend deux phases : la maturation de l'ADN viral double brin par 3'processing après la séquence consensus CA. La seconde est l'intégration à proprement parler. Dans le cas des *spumavirus*, l'analyse des extrémités LTR des provirus intégrés et des formes circulaires à 2 LTR a révélé la présence d'une asymétrie de cette étape de 3'processing. Cette dernière a été reproduite *in vitro*. L'IN des *spumavirus* ne réalise le 3'processing qu'au niveau du LTR en 3' de l'ADN viral. Cette hémimaturation des extrémités est suivie ensuite par l'intégration du génome viral au sein du génome de la cellule hôte. Le mécanisme utilisé rappelle celui utilisé par le rétrotransposon Ty de levure, où les réactions de clivage des extrémités n'ont pas lieu sur les deux brins du rétrotransposon avant son intégration ((Su and Brown, 1997); (Mules *et al.*, 1998)). L'intégration des *spumavirus* suit donc un mécanisme intermédiaire.

Les mécanismes de transcription post-intégrative des *spumavirus* ont révélé que cette étape montre elle aussi des différences avec celle des orthoretrovirus. Deux points apparaissent comme fondamentalement différents, premièrement la présence d'un promoteur fonctionnel interne au virus (Lochelt *et al.*, 1993). Deuxièmement, les *spumavirus* utilisent d'un transcrit splicé pour l'expression de la protéine Pol alors que chez les orthoretrovirus la protéine Pol provient d'un ARNm génomique ((Enssle *et al.*, 1996); (Lochelt and Flugel, 1996); (Yu *et al.*, 1996)). Il n'y a donc pas de polyprotéine Gag-Pol chez les *spumavirus* (cf. figure 15).

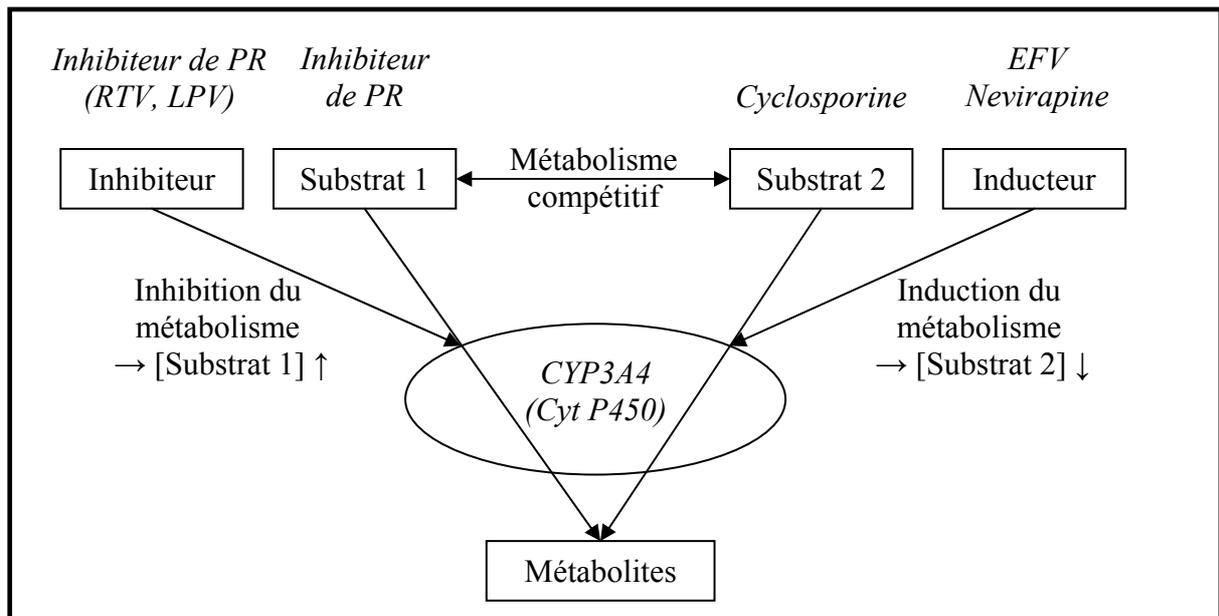
## **I.D. Les traitements anti-rétroviraux actuels**

Les médicaments anti-rétroviraux actuels ciblent quatre étapes du cycle de réplication du VIH-1 que sont la transcription inverse de l'ARN génomique en ADN viral double brin, l'intégration de l'ADN viral au sein de l'ADN de la cellule hôte, la maturation du virion, et la fusion de l'enveloppe virale avec la membrane cellulaire. Les composés actifs approuvés par la FDA (Food and Drug Administration) sont au nombre de 27 subdivisés en cinq classes. Différentes combinaisons de ces drogues entrent dans la composition des traitements anti-rétroviraux hautement actifs (HAART) qui constituent la majorité des thérapies anti-VIH utilisées actuellement.

Les inhibiteurs de la RT analogues des bases nucléosidiques / nucléotidiques bloquent la synthèse de l'ADN viral (treize médicaments dont quatre combinaisons). La molécule la plus ancienne est l'AZT (Azidothimidine). Elle a été approuvée en 1987. Cette famille d'inhibiteurs utilise la faible spécificité de reconnaissance des bases par la RT ainsi que l'absence de relecture et donc de correction lors de la transcription inverse de l'ARN en ADN viral. L'absence du groupement hydroxyl en position 3' du ribose bloque l'allongement de la chaîne d'ADN viral. La RT est ciblée par une deuxième famille de composés, les inhibiteurs de la transcriptase inverse non analogues (à ce jour, trois médicaments). Comme leur nom l'indique ces molécules sont des inhibiteurs non compétitifs des nucléosides. Ces composés se fixent sur un site différent du site actif induisant une chute de l'efficacité catalytique de l'enzyme.

En décembre 1995, le premier des neuf inhibiteurs de la protéase du VIH fut approuvé par la FDA (le Saquinavir). Ces composés inhibent de manière compétitive l'activité de PR. Cette dernière ne peut donc plus cliver la polyprotéine Gag-Pol synthétisée et la maturation des virions étant incomplète, ces particules virales néo-formées présentent une infectiosité très diminuée.

Les inhibiteurs de la PR et de la RT sont associés dans les multithérapies qui permettent de ralentir l'apparition de virus résistants aux traitements d'inhibiteurs. Ces traitements multiples mettent en évidence des actions additives, négative et/ou synergiques des médicaments, notamment par inhibition ou activation des cytochromes P450 qui vont métaboliser les drogues, en particulier CYP3A4 (cf. figure 16). La connaissance des voies métaboliques des molécules administrées permet de prévoir certaines interactions. Le choix



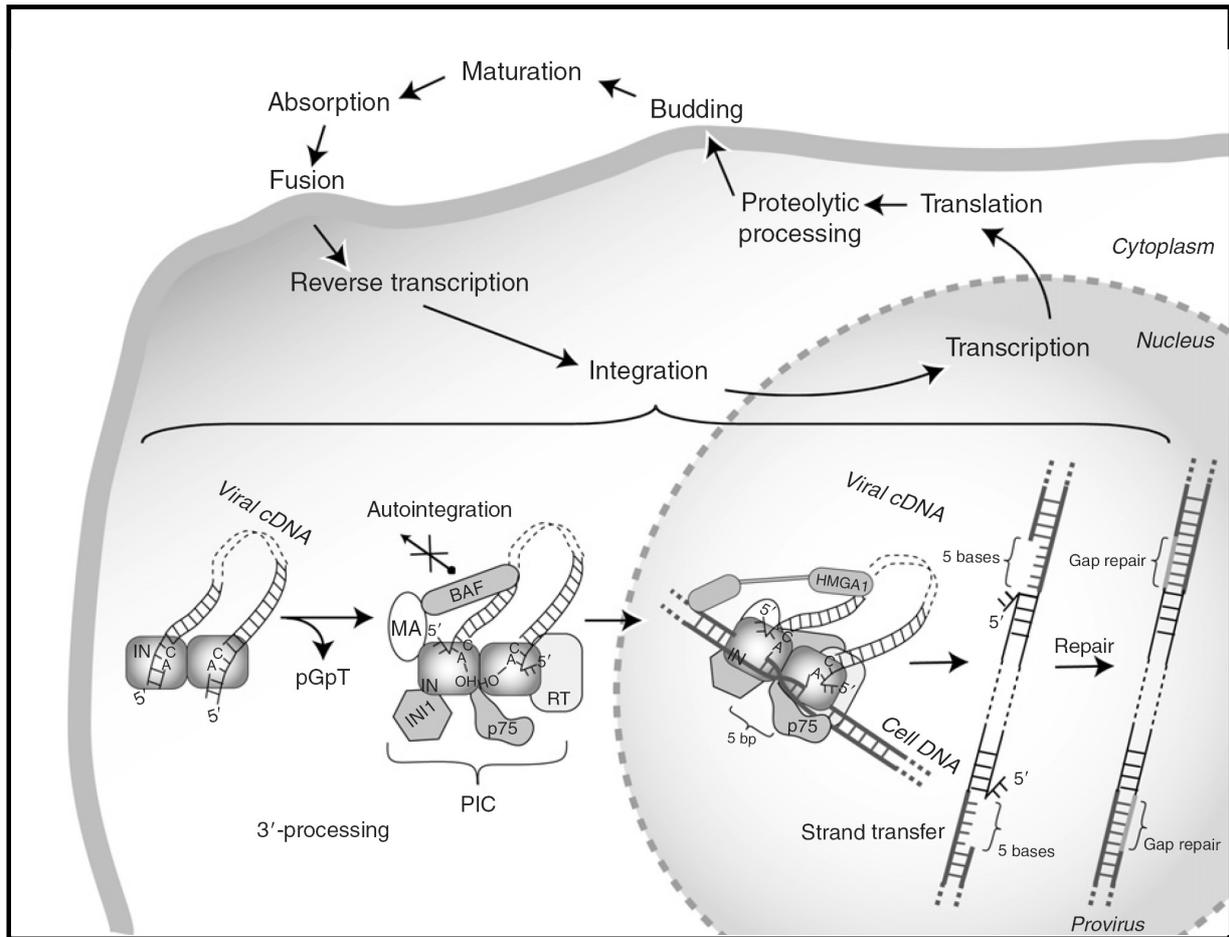
**Figure 16 : Mécanisme d'interaction métabolique drogue / drogue**

En italiques, exemples d'interaction entre des drogues vis-à-vis de CYP3A4. Les interactions peuvent être synergiques (l'antiPR Ritonavir va par son inhibition de CYP3A4 augmenter la concentration plasmatique d'autres antiPR), comme négatives (la névirapine (NNRTI) est un inducteur de CYP3A4, il va augmenter le métabolisme des antiPR). Le choix des associations des traitements est donc primordial. Exemple, la cyclosporine (anti fongique) est un compétiteur des antiPR pour CYP3A4, le métabolisme des deux molécules sera diminué, les concentrations administrées doivent être réajustées pour éviter d'atteindre des concentrations toxiques.

de la combinaison des drogues administrées est donc primordial. Il est possible alors de diminuer les doses administrées aux patients et ainsi minimiser les effets secondaires associés. Cependant, la plasticité du génome viral due à l'absence de relecture de la RT lui permet d'échapper à ces inhibiteurs combinés.

Dans le but de diversifier les cibles, deux autres étapes furent étudiées : la fusion du virion avec la cellule hôte et l'intégration du génome viral. Actuellement, il existe un inhibiteur de fusion approuvé par la FDA depuis 2003, le T-20, administré généralement lorsque les multithérapies deviennent inefficaces car le mode d'injection en sous-cutané est responsable d'effets secondaires marqués. Enfin la dernière molécule approuvée en 2007 est le Raltegravir, un inhibiteur de l'IN. Son mode d'action sera développé dans la partie consacrée aux composés anti-IN.

Toute cette pharmacopée permet de lutter à plus ou moins longue échéance contre l'apparition de virus multirésistants. Cependant, le virus de par ses propriétés de latence lorsqu'il devient résistant à un traitement, le reste quasi définitivement. De plus, il faut noter également l'existence du phénomène de résistance croisée. Ce dernier correspond à l'acquisition de mutations acquises pour résister à un inhibiteur qui lui permettent de résister à un autre inhibiteur sans que ce dernier n'ait jamais été administré au patient. Une diversification des cibles et des modes d'action des médicaments permet donc de limiter au maximum ce phénomène de résistance croisée.



**Figure 17 : L'intégration de l'ADN du VIH-1 in vivo**

Les différentes étapes du cycle de multiplication du VIH-1 avec en détail, les étapes catalysées par l'IN: le 3'processing et le transfert de brin avec les différents partenaires cellulaires et viraux impliqués (p75 : LEDGF/p75). *D'après (Semenova et al., 2008)*

## ***II. L'intégrase du VIH, sa structure, ses fonctions***

L'intégrase est la troisième protéine douée d'activité enzymatique codée par la séquence pol du génome du VIH-1. Cette protéine de la famille des polynucléotidyl transférase de 32 kDa possède une structure qui comprend trois domaines fonctionnels : le domaine N-terminal (résidus 1 à 49), le core catalytique (résidus 50 à 212) et le domaine C-terminal (résidus 213 à 288) (Engelman *et al.*, 1993).

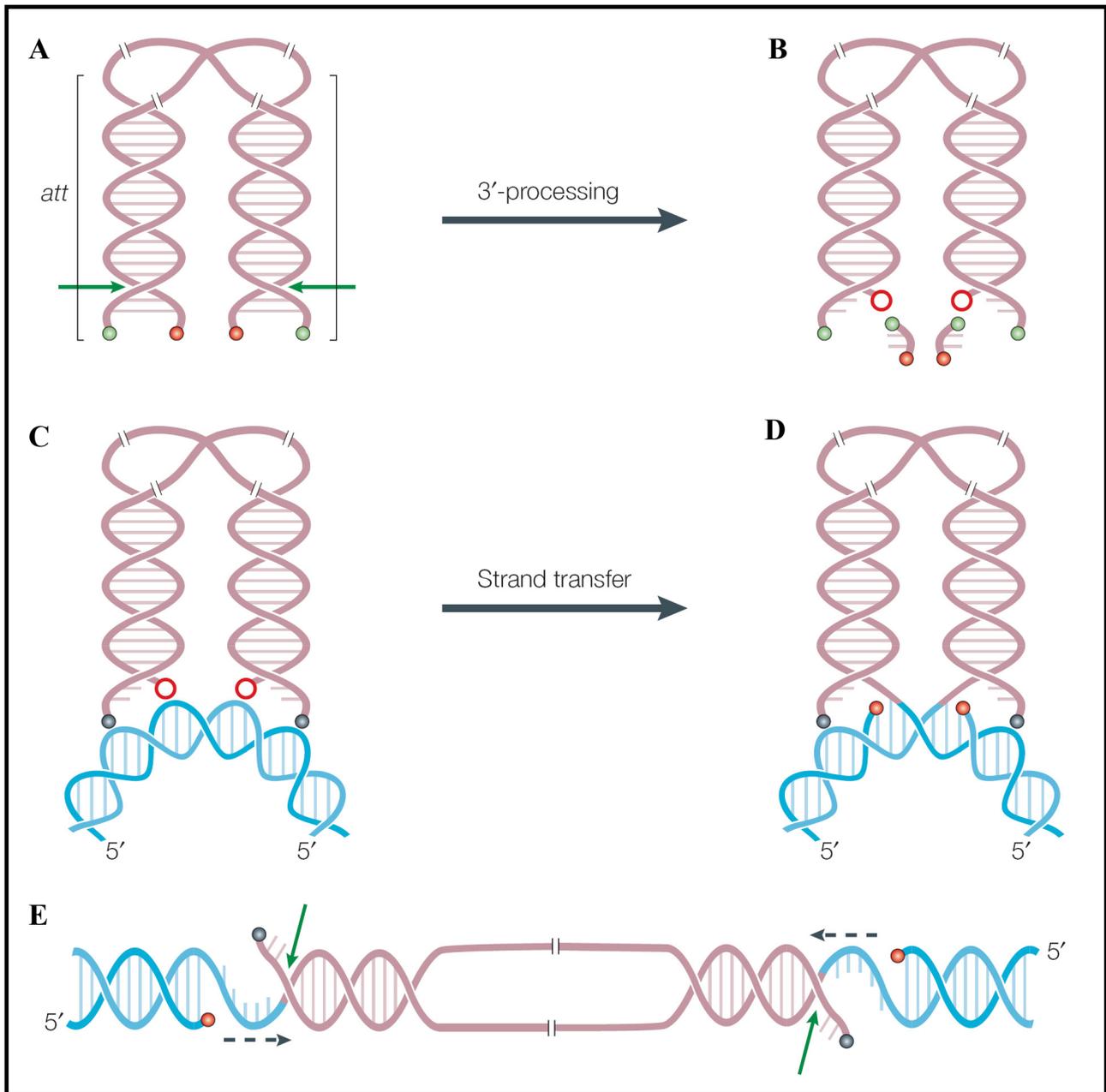
### **II.A. Les activités catalytiques de l'intégrase du VIH-1**

L'IN est l'enzyme rétrovirale qui réalise l'intégration du génome viral au sein du génome de la cellule hôte. Cette étape est un passage obligé du cycle de réplication virale et n'a lieu qu'une seule fois au cours d'un cycle répliatif.

#### ***II.A.1. Les activités de l'intégrase in vivo***

Le processus général de l'intégration covalente du génome rétroviral comprend deux étapes distinctes aussi bien spatialement que temporellement (cf. figure 17). L'IN est liée à une séquence d'attachement dite « att » située aux deux extrémités de l'ADN viral au sein des LTR. Dans un premier temps, l'IN catalyse une coupure endonucléolytique nommée 3'-processing au niveau des deux extrémités 3'-OH de l'ADN viral après un dinucléotide CA très conservé chez les rétrovirus (cf. figure 18 A et B). Cette réaction cytoplasmique est une attaque nucléophile par une molécule d'eau, suivie d'une libération dans le milieu d'un dinucléotide de séquence GT dans le cas du VIH-1 et AT dans le cas de PFV-1. L'unité catalytiquement active de cette étape n'est pas encore déterminée de manière certaine. L'ADN viral possède alors des extrémités cohésives avec les bases CA 3'-OH. Cet ADN est l'intermédiaire maturé qui va servir de substrat au double transfert de brin de l'intégration concertée.

La dernière étape se passant dans le noyau, l'IN et tout le Complexe de Préintégration (CPI) doivent traverser la membrane nucléaire au niveau des pores. Le rôle de l'IN dans le passage du pore nucléaire reste encore sujet à discussion. En effet, ce complexe nucléoprotéique d'une taille deux fois supérieure à celle des pores nucléaires, dont le diamètre intérieur est d'environ 48 nm, ne peut passer cette barrière de manière passive. Pour le



**Figure 18 : Réactions catalysées par l'IN in vivo**

Les extrémités 3'-OH sont en rouge et les 5' sont en vert

**A** : ADN viral, détail de la séquence att. La flèche verte montre la coupure endonucléolytique (3'-processing) au niveau de la séquence consensus CAGT.

**B** : Le 3'-processing effectué, l'ADN possède alors deux nouvelles extrémités 3'-OH. Le dinucléotide GT est libéré.

**C** : L'IN va, après translocation dans le noyau, rapprocher l'ADN viral et l'ADN cible.

**D** : L'intégration des deux extrémités 3'-OH néo-formées est dite concertée car elle se fait simultanément sur les deux brins d'ADN cible à 5 pb d'écart.

**E** : Les deux bases non appariées de l'ADN viral sont excisées, l'ADN sera réparé par la machinerie cellulaire.

*D'après (Pommier et al., 2005)*

moment de nombreuses hypothétiques NLS ainsi que des séquences d'adressage au noyau non conventionnelles ont été identifiées que se soit au niveau de l'IN ou de MA. Différents interactants cellulaires du CPI ont également été annoncés comme importants dans l'import nucléaire. Des membres de la famille de l'importine  $\alpha$  ou LEDGF/p75 ont été mis en évidence par différentes techniques (double hybride, immunoprécipitation,...), mais leur importance dans l'import nucléaire a été pour chacun d'eux remis en question (revue : (Suzuki and Craigie, 2007)). De ce fait, le mécanisme exact et donc le rôle de l'IN dans cette translocation au travers de la membrane nucléaire reste encore non défini (cf. figure 17), mais la karyophilie globale du CPI permet au VIH-1 de se multiplier dans des cellules sans qu'elles entrent en division comme les macrophages.

Le noyau atteint, l'IN peut alors réaliser de manière concertée pour les deux extrémités processées de l'ADN viral l'intégration du génome viral au sein du génome cellulaire. Pour cela, les deux ADN, viral et cible, doivent interagir de manière stable. Si l'IN seule suffit, d'autres partenaires cellulaires permettent de moduler l'intégration. L'interaction de l'IN avec LEDGF/p75 ((Maertens *et al.*, 2003) ; (Llano *et al.*, 2006)) permet d'augmenter l'efficacité de l'intégration du provirus, mais également d'avoir une sélection plus ou moins marquée du site de l'intégration (cf. figure 17). Des études récentes utilisant notamment les siRNA ont montré l'implication de l'emerine (une protéine présente au niveau de la face interne de la membrane nucléaire) et de BAF dans cet ancrage du CPI à l'ADN cellulaire (Jacque and Stevenson, 2006). Cependant, l'implication de l'emerine pourrait déjà être remise en cause notamment sur un modèle de fibroblastes embryonnaires murins knock-out pour l'emerine dans lesquels le VIH-1 se réplique normalement (Shun *et al.*, 2007a). Il est donc difficile encore maintenant de dire précisément quels sont les déterminants et les partenaires de l'IN au cours de cette étape qui précède l'intégration.

Cette réaction est une attaque nucléophile des extrémités 3'-OH de l'ADN viral processé sur deux 5'-phosphates situés sur des brins opposés et séparés par 5 paires de bases, ce qui correspond à un écart d'environ 15 Å si l'on considère la forme hélicoïdale B de l'ADN (Varmus and Paul, 1997). Pour réaliser l'intégration concertée, il faut donc que les deux sites actifs de l'unité fonctionnelle soient séparés d'environ 15 Å (cf. figure 18 C et D). La distance entre les sites actifs dans la maille cristallographique de chacune des structures de domaines de l'IN contenant le site catalytique (IN<sup>1-212</sup>, IN<sup>55-212</sup> et IN<sup>55-288</sup>) est d'environ 30 Å et l'IN est alors sous forme dimérique. Pour trouver une distance entre deux sites actifs de

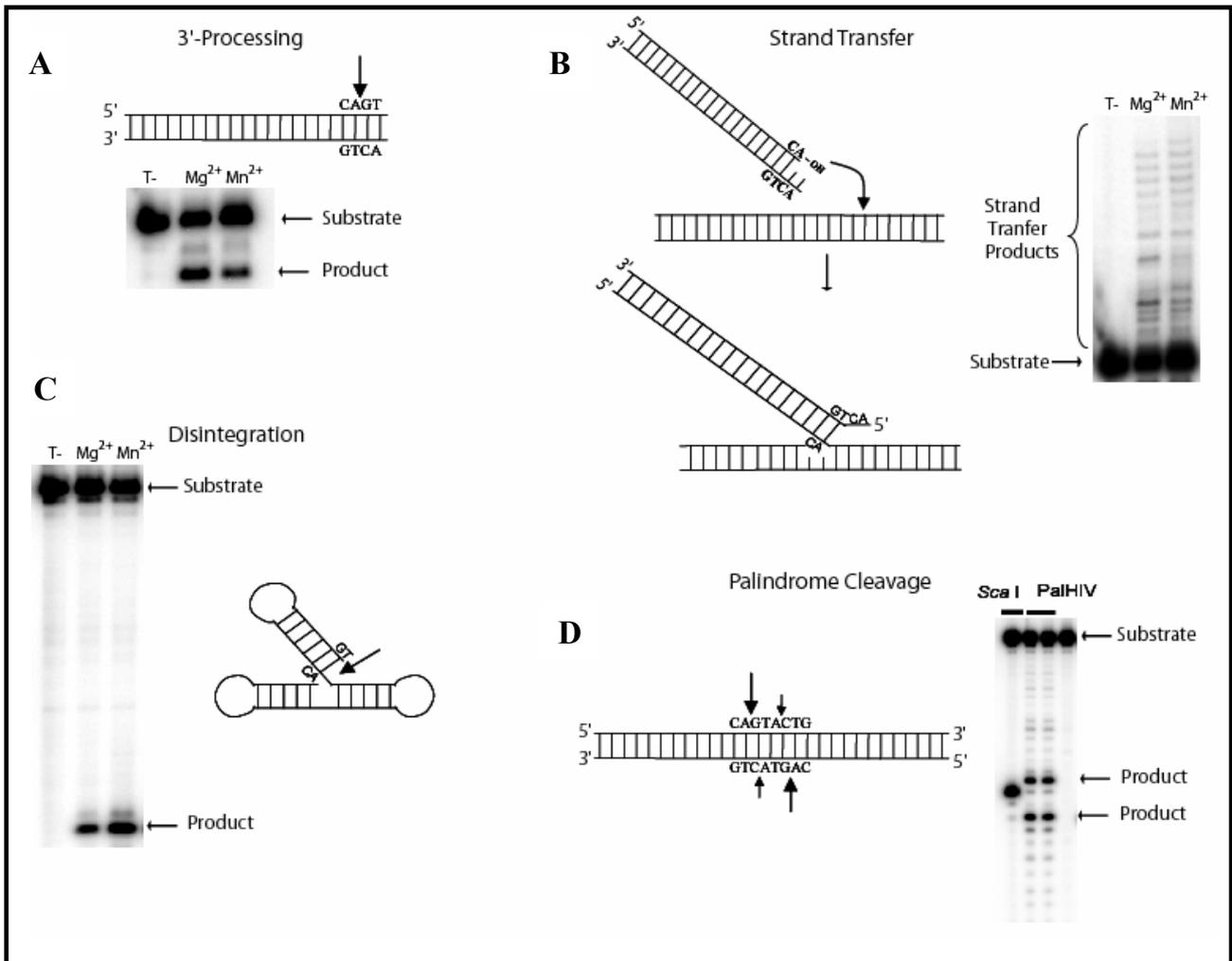
l'ordre de 15 Å, il faut que l'IN soit dans un état multimérique d'ordre supérieur à deux. Différents modèles du complexe réalisant l'intégration concertée ont été suggérés, ils font tous intervenir un tétramère d'IN ((Podtelezhnikov *et al.*, 2003) ; (Wielens *et al.*, 2005)).

Au final, dans l'ADN cellulaire, l'intégration concertée crée une cassure double brin qui aboutit à des interruptions de la structure nucléoprotéique de la chromatine, à deux courtes séquences non appariées simples brins flanquant l'ADN proviral et des interruptions en interne de cet ADN nouvellement intégré (cf. figure 17 et 18 E). Ces évènements sont également retrouvés lors de la réplication de l'ADN cellulaire notamment au niveau des fourches de réplication. Afin d'achever l'intégration, les deux CA non appariés sont éliminés de chaque extrémités 5' de l'ADN viral. L'espace compris entre le 5'-OH de l'ADN cellulaire et 3'-OH de l'ADN viral est ensuite réparé afin d'éviter l'entrée en apoptose de la cellule hôte. Une étude, non encore confirmée à ce jour, a mis en évidence une activité ADN polymérase de l'IN lui permettant ainsi de combler ces espaces (Acel *et al.*, 1998). Ceci étant dit, les études les plus récentes tentent à montrer que c'est la machinerie cellulaire qui est responsable de cette réparation. Actuellement, plusieurs protéines cellulaires sont susceptibles de jouer ce rôle de réparation de l'ADN.

- La voie NHEJ (Non Homologous End-Joining) représentant le mécanisme majeur de réparation des cassures doubles brins de l'ADN, est la voie la plus étudiée. Cette voie est médiée par la DNA PK qui est activée par l'extrémité cassée de l'ADN. Elle est composée d'une sous unité catalytique (DNAPK<sub>CS</sub>) et de l'hétérodimère Ku70/86 qui reconnaît l'ADN cassé. L'implication de la voie NHEJ dans cette réparation fait l'objet d'une certaine controverse ((Daniel *et al.*, 1999) ; (Baekelandt *et al.*, 2000)).
- D'autres facteurs cellulaires sont potentiellement impliquée : la Poly (ADP-ribose) polymérase 1 (PARP-1) qui est activée par les cassures simple et double brins (Ha *et al.*, 2001).
- Une voie de réparation en réponse à l'activation de la kinase ATR qui est impliquée dans la régulation des points de contrôle du cycle cellulaire (Daniel *et al.*, 2003) pourrait également être impliquée.
- Enfin, deux autres facteurs cellulaires RAD18 (facteur réparant les lésion post-répliquatives et les trans-lésions) et RAD52 (facteur de la voie de recombinaison

homologue) pourraient également jouer un rôle dans la réparation de l'ADN ((Mulder *et al.*, 2002) ; (Lau *et al.*, 2004)).

Si la séquence att terminale de l'ADN viral est primordiale pour que se réalise l'étape de 3'-processing, il y a peu de spécificité quand à la séquence d'intégration au niveau du génome de la cellule cible. Il faut cependant noter que dans le cas de VIH-1, l'intégration se fait en général au niveau de gènes transcrits par l'ARN pol II et que les régions transcriptionnellement inactives comme les télomères et centromères présentent statistiquement peu de phénomènes d'intégration. De plus, un certain consensus pour les séquences d'intégration comportant une symétrie de type palindromique a pu être mis en évidence pour différents rétrovirus dont le VIH-1 (Wu *et al.*, 2005). Il a été récemment montré que l'IN pouvait être acétylée au niveau de son domaine C-terminal par une histone acétyl-transférase. Si, *in vitro*, cette acétylation ne modifie en rien la réaction de 3'-processing, elle stimule de manière significative la réaction de transfert de brin (qui correspond à une demi-intégration). Cependant, l'acétylation de l'IN *in vitro* n'ont pas encore pu être confirmées *in vivo* ((Cereseto *et al.*, 2005) ; (Topper *et al.*, 2007)). Il ne fait aucun doute que l'intégration rétrovirale est influencée par la structure chromatinienne. Dans le cas du VIH-1, il a été montré par la technique du double hybride qu'une sous unité du complexe SWI / SNF, qui est impliqué dans les changements de structure chromatinienne, SNF5 / Ini 1 interagit avec l'IN (Kalpana *et al.*, 1994). Cependant, *in vivo*, le rôle de cette interaction lors de l'intégration n'a pas encore été démontré de manière claire. Cette hypothèse de l'importance de la structure chromatinienne dans le choix des sites d'intégration est renforcée par la mise en évidence des interactions entre les intégrases lentivirales dont celle du VIH-1 et LEDGF/p75, qui est un facteur de croissance associé à la chromatine. L'interaction avec le CPI va induire un ciblage spécifique, sans qu'il soit pour autant obligatoire, de la zone d'intégration de l'ADN viral par l'intermédiaire de LEDGF/p75. Des expériences utilisant la technique des siRNA contre LEDGF/p75 ont montré une redistribution des zones d'intégration du VIH-1 dans le génome des cellules hôtes. En présence de LEDGF/p75, les provirus sont intégrés de manière préférentielle au sein des unités transcriptionnelles actives. En son absence, la répartition des provirus intégrés est redistribuée vers les zones de type CpG et vers les promoteurs transcriptionnels, comme c'est le cas pour les autres rétrovirus (Shun *et al.*, 2007b).



**Figure 19 : Réactions catalysées par l'IN in vitro**

**A** : Réaction de 3'-processing : coupure endonucléolytique après le CA.

**B** : Réaction de transfert de brin : insertion covalente du CA-OH en interne d'un oligonucléotide.

**C** : Réaction de désintégration : l'IN clive un oligonucléotide au repliement en Y mimant le produit du transfert de brin au niveau du CA conservé.

**D** : Réaction de coupure d'une séquence de type palindromique : l'IN clive cette séquence palindromique après chacun des deux CA (coupure principale) et après le A qui suit le CAGT (coupure secondaire), à la manière d'une enzyme de restriction.

*D'après Delelis et al., revue en préparation*

## ***II.A.2. Les activités de l'intégrase observées in vitro***

Afin de mieux comprendre le mécanisme catalytique de l'IN et de trouver des inhibiteurs efficaces, différents tests d'activité catalytique de l'IN ont été mis au point *in vitro*. Ces tests miment deux réactions mises en évidence *in vivo* : le 3'-processing et le transfert de brin. L'IN recombinante est mise en présence d'ADN double brin qui miment la séquence terminale U3 ou U5 des LTR viraux (Leh *et al.*, 2000). La réaction de 3'-processing est une réaction très spécifique et séquence dépendante. Cette réaction implique la coupure endonucléolytique d'un dinucléotide GT situé après la séquence CA conservée chez les rétrovirus du brin 3' des deux extrémités des LTR (cf. figure 19 A). Une mutation dans ces bases abolit complètement toute activité de 3'-processing. Les séquences adjacentes même si elles sont importantes restent moins stringentes vis-à-vis de cette activité (Esposito and Craigie, 1998). Le 3'-processing correspond à l'attaque nucléophile d'une molécule d'eau. Cependant, d'autres nucléophiles peuvent être utilisés comme, par exemple le glycérol, mais la coupure résultante présente une spécificité réduite (Engelman *et al.*, 1991). Lorsque l'on utilise comme cation divalent le  $Mn^{2+}$ , la stringence vis-à-vis de l'agent nucléophile est également réduite. L'utilisation du  $Mg^{2+}$ , qui semble être le cofacteur biologique, permet d'obtenir une coupure plus spécifique qu'en présence de  $Mn^{2+}$ .

Lors de la réaction de 3'-processing, il est possible d'observer dans une plus faible mesure du transfert de brin que l'on peut quantifier entre 1 et 5 % (cf. figure 19 B). Au cours de cette réaction, l'agent nucléophile est l'extrémité 3'-OH du brin traité. Ces faibles pourcentages peuvent être augmentés lorsqu'il y a déjà dans le mélange réactionnel de l'ADN pré-traité (Li and Craigie, 2005). En utilisant un oligonucléotide mimant une seule extrémité de l'ADN viral, on reproduit *in vitro* une demi-intégration. L'intégration concertée complète est encore plus difficile à reproduire *in vitro*. Pour cela, il faut utiliser des oligonucléotides plus longs. Ces expériences ont permis de montrer que l'IN seule était capable de réaliser cette réaction ((Li *et al.*, 2005) ; (Li *et al.*, 2006) ; (Sinha *et al.*, 2002) ; (Sinha and Grandgenett, 2005)). Par contre, différents facteurs cellulaires et viraux sont apparus comme étant des cofacteurs permettant de stimuler ce processus d'intégration concertée comme la protéine NC (Carteau *et al.*, 1999), les facteurs cellulaires HMG I(Y) (Hindmarsh *et al.*, 1999) et LEDGF/p75 ((Cherepanov *et al.*, 2003) ; (Cherepanov, 2007)). Contrairement au transfert de brin, lors d'expériences mimant l'intégration concertée, l'utilisation d'ADN pré-traité n'augmente pas l'efficacité de la réaction, au contraire,

l'efficacité de la réaction est réduite. Ceci démontre que la réaction d'intégration concertée nécessite une conformation de l'IN qui va être acquise ou facilitée par la réaction de 3'-processing (Li *et al.*, 2005).

Une troisième réaction est également réalisée *in vitro* : la désintégration. Cette réaction correspond à la réaction inverse du transfert de brin (Chow *et al.*, 1992). L'IN va cliver un ADN replié en une forme caractéristique de Y mimant ainsi une demi-intégration au niveau d'une séquence comprenant le CA terminal du brin 3' qui aurait été processé avant le transfert de brin (cf. figure 19 C). La structure minimale requise pour avoir une réaction de désintégration comprend uniquement 5 paires de bases hybridées TAGCA qui correspondent à l'extrémité du LTR processé (Chow and Brown, 1994). Historiquement, de part la faible stringence requise pour son bon déroulement, la désintégration a joué un rôle essentiel dans la localisation du domaine catalytique de l'IN ((Engelman and Craigie, 1992) ; (Vink *et al.*, 1993)). Cette réaction est encore aujourd'hui intéressante car elle peut être catalysée, contrairement au 3'-processing et au transfert de brin, par des intégrases tronquées d'un de leur domaine C- ou N-terminal (IN<sup>1-212</sup> et IN<sup>55-288</sup>) ((Leh *et al.*, 2000), (Laboulais *et al.*, 2001)) ou le core catalytique seul (IN<sup>55-212</sup>). La désintégration comme les autres activités *in vitro* nécessite un cation divalent qui agit comme un cofacteur de l'IN. Le cofacteur préférentiel de cette réaction est le Mn<sup>2+</sup>, mais elle peut également être réalisée en présence de Mg<sup>2+</sup>. Les IN tronquées ne réalisent cette réaction qu'en présence de Mn<sup>2+</sup>. Cette activité n'a encore jamais été mise en évidence *in vivo*, mais elle est utilisée pour mettre en évidence le type d'inhibition compétitive de certains inhibiteurs de l'IN.

Une quatrième activité de l'IN du VIH-1 a été mise en évidence *in vitro* (Delelis *et al.*, 2007). Cette nouvelle activité, également observée pour l'IN de PFV-1 (Delelis *et al.*, 2005) est une activité de coupure endonucléolytique au niveau d'une séquence spécifique située en interne. Cette réaction très spécifique est possible à partir d'oligonucléotides ou de plasmides comportant la séquence palindromique que l'on retrouve au niveau de la jonction LTR - LTR des cercles à 2-LTR. La coupure a lieu de manière symétrique sur les deux brins antiparallèles avec une majorité des clivages réalisés après le dinucléotide CA qui correspond exactement à celui du 3'-processing. Une deuxième coupure est également identifiable, mais dans une moindre mesure, après l'adénine du TA qui suit directement le CA dans le sens 5'-3' (cf. figure 19 D). La cinétique de cette coupure, catalysée également sur un plasmide mimant un cercle à 2-LTR par l'IN, est comparable à celle observée pour le 3'processing, elle est réalisée

par une unité fonctionnelle d'IN tétramérique et en présence de l'un ou l'autre des cations divalents possibles : le  $Mg^{2+}$  et le  $Mn^{2+}$ .

Comme cela a déjà été explicité précédemment, la présence de cations divalents est nécessaire pour l'activité catalytique de l'IN. Les cations servant de cofacteurs utilisés *in vitro* sont soit le  $Mn^{2+}$ , soit le  $Mg^{2+}$ . Actuellement, toutes les données expérimentales laissent à penser que le cofacteur utilisé préférentiellement *in vivo* est le  $Mg^{2+}$ . En effet, le  $Mg^{2+}$  est, dans les cellules, un million de fois plus abondant que le  $Mn^{2+}$ , ce qui statistiquement lui donne une probabilité plus grande d'être disponible pour l'IN lors de l'acquisition de sa structure tertiaire. De plus, en présence de ce cation, l'IN présente, *in vitro*, une spécificité plus grande lors de la catalyse de la coupure endonucléolytique. En effet, en présence de  $Mn^{2+}$ , l'IN montre une plus grande propension à réaliser des coupures endonucléolytiques non spécifiques, c'est-à-dire aux positions -1, -3, -4 etc.... alors qu'en présence de  $Mg^{2+}$ , la coupure en position -2 est largement prépondérante (Engelman and Craigie, 1995). Il y a également en présence de  $Mn^{2+}$ , une tolérance plus grande vis-à-vis de la séquence terminale de l'ADN viral qu'en  $Mg^{2+}$  (Esposito *et al.*, 1998) (cf. figure 20 A). Enfin, en présence de  $Mn^{2+}$ , différentes mutations de l'IN restent silencieuses alors qu'elles s'expriment en présence de  $Mg^{2+}$ . Par exemple, des mutations situées au niveau du doigt de zinc du domaine N-terminal qui se révèlent être délétères *in vivo* affecte les activités de 3'-processing et de transfert de brin *in vitro* en présence de  $Mg^{2+}$ , mais pas en présence de  $Mn^{2+}$  (Wiskerchen and Muesing, 1995). Les deux cofacteurs cationiques présentent également des différences notables lors de tests d'activité en présence de drogues potentiellement inhibitrices de l'IN. D'une part, les  $IC_{50}$  résultants peuvent être différents selon le cation ajouté ((Molteni *et al.*, 2000) ; (Neamati *et al.*, 2002)) et d'autre part, les mutations de résistance développées en présence d'une drogue ont tendance à s'exprimer en  $Mg^{2+}$  et alors qu'elles restent sans effets en  $Mn^{2+}$  (Grobler *et al.*, 2002) (cf. figure 20 B). De ce fait, l'action inhibitrice d'une molécule sur l'IN en présence de  $Mg^{2+}$  est généralement plus prédictif de ce qui va être observé *in vivo*.

### ***II.A.3. Modélisation du mécanisme catalytique de l'intégrase du VIH-1***

Le processus d'intégration est caractérisé par une cinétique extrêmement lente que se soit *in vitro* ou *in vivo*. Les phénomènes impliquant une telle cinétique peuvent s'expliquer par des événements lents précédant l'étape de catalyse chimique. En effet, la coupure endonucléolytique ne peut pas présenter une telle cinétique. Nous allons nous attacher dans

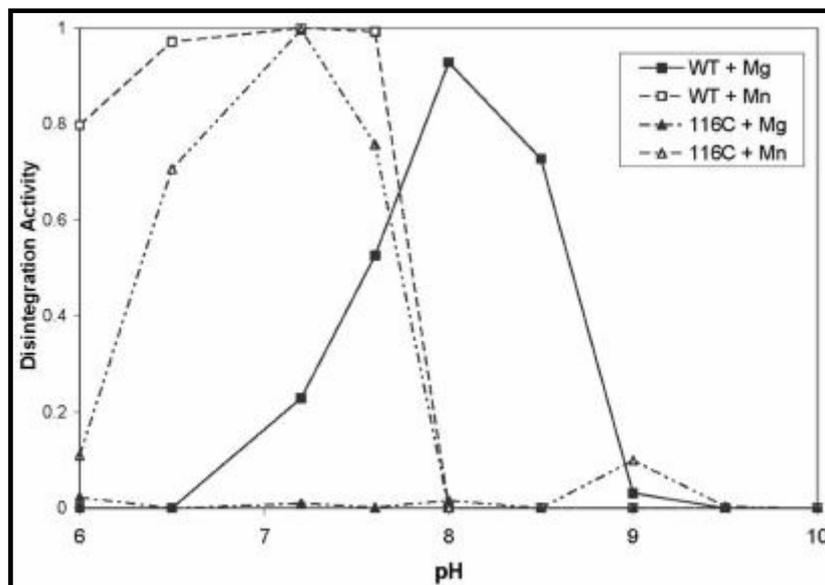


cette partie à décrire les événements purement chimiques qui ont lieu durant les réactions catalysées par l'IN du VIH-1.

### II.A.3.a. Les réactions catalysées par l'IN sont de type $S_N2$

Quelque soit le modèle utilisé, les deux réactions successives que sont le 3'-processing et le transfert de brin sont réalisés par le même site actif. Le mécanisme impliqué est une transestérification réalisée en une seule étape (Engelman *et al.*, 1991). Gerton *et al.* (1999) (Gerton *et al.*, 1999) ont montré que le mécanisme chimique suivi est une substitution nucléophile de type 2 ( $S_N2$ ). Ils ont utilisé des oligonucléotides pour lesquels le phosphate clivé est remplacé par un phosphorothioate. Ce changement permet d'observer les formes énantiomériques du phosphate prochiral. Les résultats montrent que l'IN, lors du 3'-processing et du transfert de brin, clive les substrats dont la configuration absolue du phosphorothioate est Rp. Les produits de la réaction ont une configuration absolue Sp. Ce résultat permet de conclure quant à un mécanisme de  $S_N2$  en une seule étape avec une inversion de configuration absolue dite de Walden. Lors de la réaction de désintégration, le substrat préférentiel est de configuration absolue Sp pour devenir Rp après la réaction. Ces résultats confirment l'hypothèse que la désintégration est la réaction inverse du transfert de brin.

A partir de ces résultats, Gerton *et al.* proposent un modèle qui prend en compte que l'IN n'utilise qu'un seul site actif pour catalyser ces trois réactions. Ils tiennent également compte du fait que la paire de base CA/TG est cruciale pour les trois réactions et qu'elle forme donc avec l'IN des interactions spécifiques au niveau du site actif. Enfin, lors de l'état de transition de chacune des réactions, l'atome de phosphore adopte une géométrie bipyramidale à base triangulaire caractéristique des  $S_N2$  (cf. figure 21). Dans cette géométrie, le nucléophile et le groupe partant occupent les deux positions apicales. Les positions équatoriales seront occupées par les autres substituants du phosphore. Le fait que l'IN n'agisse clairement que sur un seul énantiomère pour chaque type de réaction, montre qu'elle interagit avec un seul des deux oxygènes du phosphate. Au niveau macroscopique, la désintégration correspond à l'inverse du transfert de brin. Ce modèle montrerait qu'au niveau moléculaire, ce serait aussi le cas. Cependant, ce modèle présente un certain nombre de défauts puisqu'il ne tient pas compte de l'introduction du soufre à la place de l'oxygène. En effet, si au niveau stérique il n'y a pas beaucoup de changements, au niveau purement



**Figure 22 : Activité de l'IN en fonction du pH et du cation divalent**

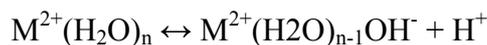
*D'après (Gao et al., 2004)*

En noir, l'IN est en présence de  $Mg^{2+}$ , en blanc de  $Mn^{2+}$ . les carrés sont pour l'IN sauvage, les triangles pour l'IN mutée D116C.

électronique, le soufre introduit une différence de charges coulombiennes à l'intérieur du phosphothioate qui n'est pas présente au niveau du phosphate. De plus, si l'énantiomère Sp a peu d'effets sur la conformation structurale de l'ADN, l'énantiomère Rp a tendance à élargir le petit sillon de l'ADN de forme B et cette modification conformationnelle de l'ADN qui est connu pour favoriser les interactions protéines ADN (Jaroszewski *et al.*, 1992).

### II.A.3.b. Dissection de la catalyse à l'échelle atomique

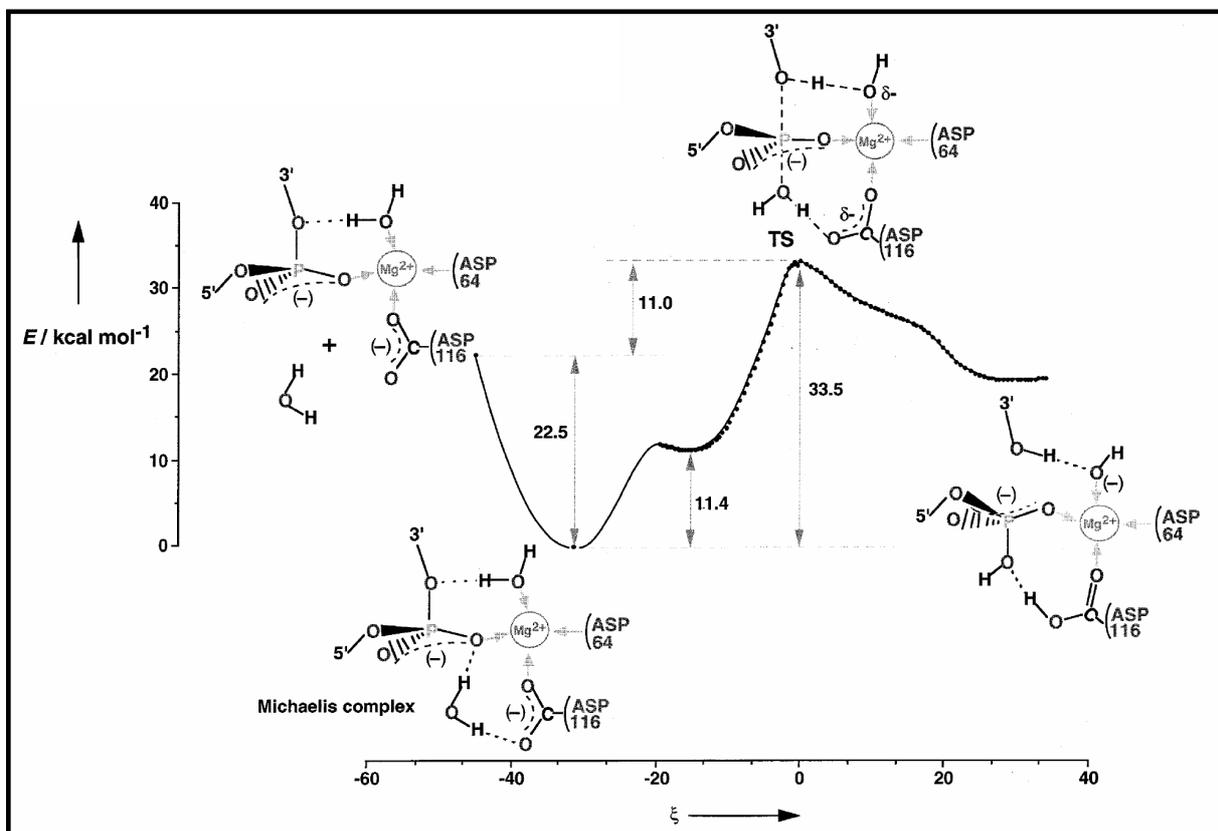
Si le modèle proposé par Gerton et al. considère que l'agent nucléophile est une molécule d'eau, le véritable attaquant d'une  $S_N2$  est un ion  $OH^-$  dans le cas du 3'-processing et l'extrémité 3'O<sup>-</sup> de l'adénine dans celui du transfert de brin. Cet agent nucléophile peut être formé à partir de la sphère d'hydratation du cation divalent selon la réaction chimique suivante :



Le rapport entre les deux formes dépend directement du pH du milieu réactionnel et est donné par la formule suivante :

$$\log ([M^{2+}(H_2O)_{n-1}OH^-] / [M^{2+}(H_2O)_n]) = pH - pK1$$

Le pK1 dépend de chaque cation divalent considéré. Le pH donnera donc le nombre d'attaquant nucléophile, sachant que celui-ci sera au maximum égal au nombre de cations divalents présents. L'activité est également limitée par les interactions IN / ADN. Beaucoup d'entre elles sont régies par des interactions ioniques mettant en jeu des lysines. Leur neutralisation à pH basiques (au-delà de 9) explique pourquoi les intégrases rétrovirales ne sont pas actives au-delà de ces pH. Cette observation a été faite sur l'IN d'ASV (Lubkowski *et al.*, 1998). Dans le cas de l'IN du VIH-1, l'activité chute avant d'atteindre ces valeurs. Ceci montre la présence d'un autre facteur limitant. Si l'on considère le cation divalent mis en présence dans le tampon,  $Mn^{2+}$  ou  $Mg^{2+}$ , leur pK1 sont respectivement de 10,46 et de 11,40. L'IN en leur présence présente un décalage du pH optimal pour l'activité de 7 à 8 (Gao *et al.*, 2004) (cf. figure 22). Le décalage des deux pK1 est donc retrouvé lorsque l'on compare les pH optimum d'activité. Ces données confirment l'importance et le rôle du cation divalent dans la formation de l'ion hydroxyde.



**Figure 23 : Diagramme énergétique de la réaction de  $S_N2$  catalysée par l'IN**

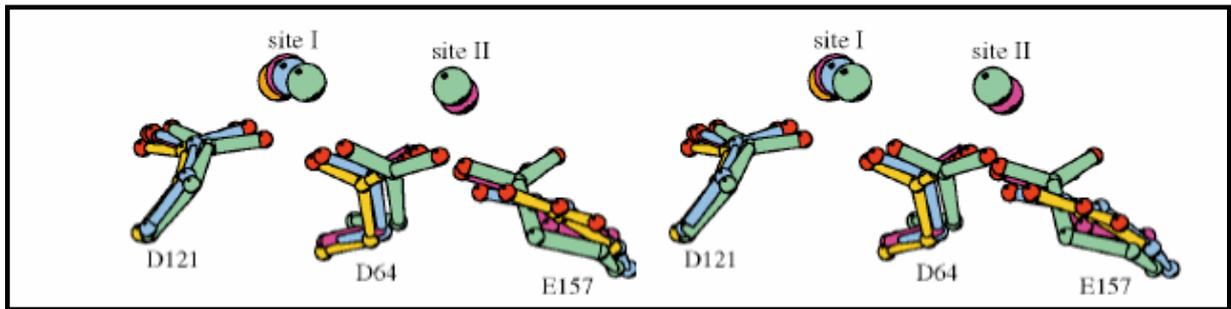
Le complexe de Michaelis (l'état de transition) fait intervenir les deux Asp 64 et 116 qui coordonnent le  $Mg^{2+}$ , l'Asp 116 interagit en plus avec une molécule d'eau, il va accélérer sa déprotonation et donc faire chuter la barrière énergétique de 70,3 kcal/mol à 33,5 kcal/mol. Au final l'Asp 116 est protonné.

*D'après (Bernardi et al., 2002)*

L'analyse des séquences des intégrases rétrovirales suggère que certains résidus pourraient jouer un rôle dans le mécanisme catalytique des intégrase, notamment en cédant un proton. Selon l'IN considérée, ASV, VIH-1 ou MLV, on observe des pH optimums d'activité en présence de  $Mn^{2+}$  différents qui sont respectivement de 8,5 (Lubkowski *et al.*, 1998), 7,0 (Gao *et al.*, 2004) et 6,2 (Jonsson *et al.*, 1993). Le résidu impliqué serait le résidu correspondant au résidu 114 dans la séquence primaire de l'IN du VIH-1. Dans le cas d'ASV, c'est une lysine, pour HIV-1, c'est une histidine et dans le cas de MLV, c'est une glycine, pour cette IN, ce serait le solvant qui donnerai un proton, d'où un décalage du pH optimum vers des valeurs acides, qui suivent le décalage des pKa des groupements donneurs potentiels de protons. D'autres résidus peuvent être impliqués directement dans le mécanisme catalytique. La chute rapide de l'activité des intégrases lorsque le pH s'acidifie est symptomatique d'une part d'une diminution de l'ionisation de la sphère d'hydratation du cation divalent, et d'autre part d'une protonation progressive des groupes carboxyliques de la triade catalytique, qui ne peuvent alors plus jouer leur rôle, notamment dans la coordination du cation divalent.

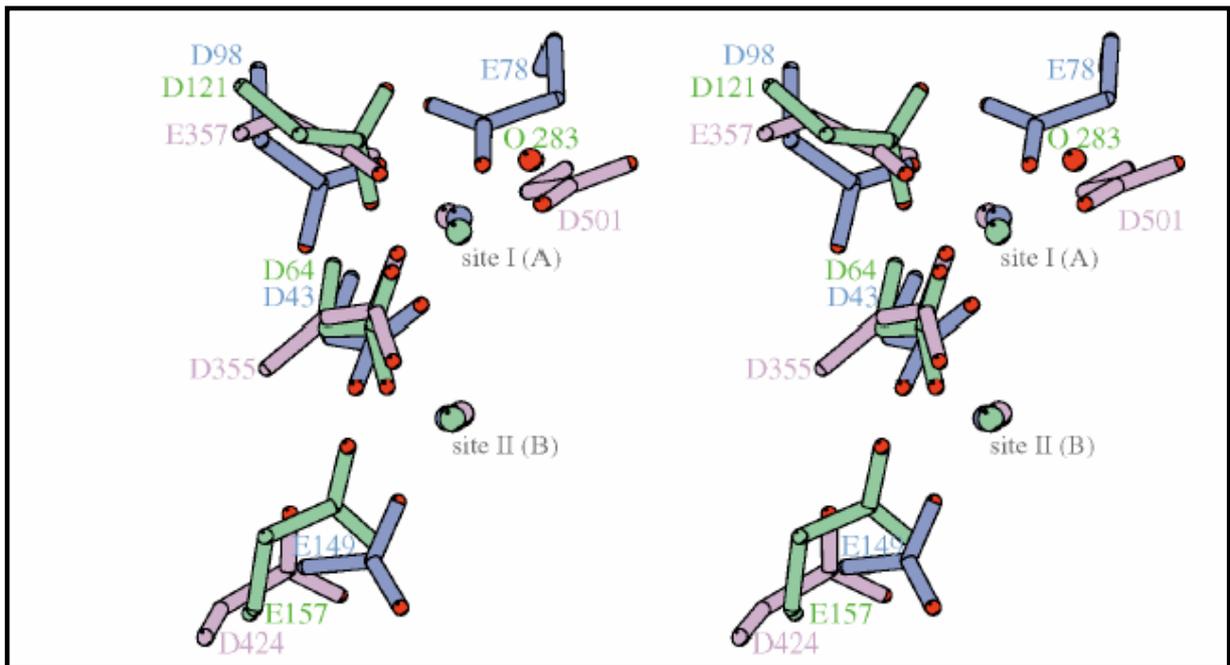
Lors de la mutation de l'Asp 116 en cystéine, résidu au pKa plus élevé (on passe de 4,4 à 8,5), la diminution de l'activité, que l'on observe lors de l'acidification du pH, est plus rapide (Gao *et al.*, 2004) (cf. figure 22). Cette observation implique que ce résidu recevrait un proton au cours de la réaction chimique. Le mécanisme proposé par Bernardi *et al.* (Bernardi *et al.*, 2002) s'appuie sur cette observation et utilise l'Asp 116 pour polariser une molécule d'eau qui va alors par perte d'un proton devenir l'attaquant nucléophile ( $OH^-$ ). C'est une autre molécule d'eau, qui appartient à la sphère d'hydratation du cation divalent, qui va fournir le proton de l'extrémité 3'-O<sup>-</sup> du nucléotide partant, qui devient 3'-OH, et améliore ainsi son pouvoir partant. Ce mécanisme permet un abaissement de la barrière énergétique de la réaction de 70,3 kcal/mol à 33,5 kcal/mol en présence d'IN (cf. figure 23).

Dans les mécanismes précédents, un seul cation divalent est présent au niveau du site actif. Ceci est en accord avec les structures cristallographiques de l'IN du VIH-1 dans lesquelles le cation est le centre d'un complexe de coordination comprenant quatre molécules d'eau et les deux groupements carboxylates des Asp 64 et 116 de la triade catalytique. En considérant la famille des polynucléotidyl-transférases qui comprend notamment les IN et d'autres enzymes comme le fragment de Klenow de l'ADN polymérase I, le nombre de cations impliqués dans la catalyse ne serait pas de un mais de deux. En effet, la résolution de



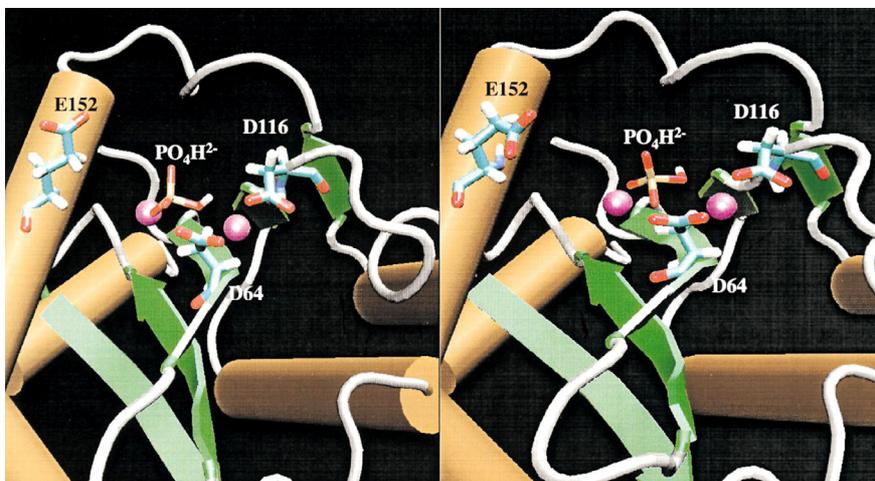
**Figure 24 :** Vue en stéréo des 2 sites de fixation des cations divalents de l'IN d'ASV

Le site I serait occupé par le  $Mg^{2+}$  et le site II par un  $Zn^{2+}$ . D'après (Bujacz et al., 1997)



**Figure 25 :** Comparaison des sites actifs de l'IN d'ASV, la RNase H et du fragment de Klenow

Vue stéréo des sites actifs de, en vert l'IN d'ASV, en bleu la RNase H du VIH-1, en rose le fragment de Klenow, positionnement identique des deux sites des cations divalents. D'après (Bujacz et al., 1997)



**Figure 26 :**  
Configuration du site actif de l'IN du VIH-1 avec 2  $Mg^{2+}$  au site actif.

Gauche : le site actif et le 2<sup>nd</sup>  $Mg^{2+}$  au site II.

A droite : la chaîne latérale de E152 est déplacée de manière à ce qu'elle coordonne le 2<sup>nd</sup>  $Mg^{2+}$ . D'après (Lins et al., 2000)

la structure du fragment de Klenow (Beese and Steitz, 1991) et de celle du core catalytique de l'IN d'ASV montrent la présence d'un deuxième positionnement possible pour un cation divalent au niveau du site actif (cf. figure 24 et 25). Par analogie, ce deuxième  $Mg^{2+}$  serait coordonné par le premier et le troisième résidu de la triade catalytique (Asp 64 et Glu 152). Pour que ce  $Mg^{2+}$  soit coordonné, la chaîne latérale du Glu 152 doit être redirigée vers le site actif par rapport à son positionnement observé dans les différentes structures cristallographiques obtenues ((Maignan *et al.*, 1998) ; (Goldgur *et al.*, 1998)) (cf. figure 26). Ce deuxième cation possède une stabilité plus faible que le premier et les différentes modélisations le montrent stabilisé s'il est en présence du phosphate présent dans la structure résolue de Goldgur et al. (Lins *et al.*, 2000). De ce fait, sa position pourrait être stabilisée par la présence de l'ADN viral ou cible. Ces deux cations auraient des rôles différents : le premier, comme dans l'hypothèse d'un mécanisme à un  $Mg^{2+}$ , activerait une molécule d'eau pour qu'elle réalise l'attaque nucléophile, le second  $Mg^{2+}$  permettrait de stabiliser l'état de transition du phosphate pentacoordiné.

Le nombre de cations divalents au site actif reste encore indéterminé et étant donné le rôle qu'ils pourraient jouer dans des changements conformationnels du site actif que se soit en présence ou non d'ADN viral comme cible (Asante-Appiah *et al.*, 1998). Cela représente un frein pour l'amélioration des drogues inhibitrices de l'IN actuellement conçues et qui possèdent la capacité à chélater le  $Mg^{2+}$  (Savarino, 2007).

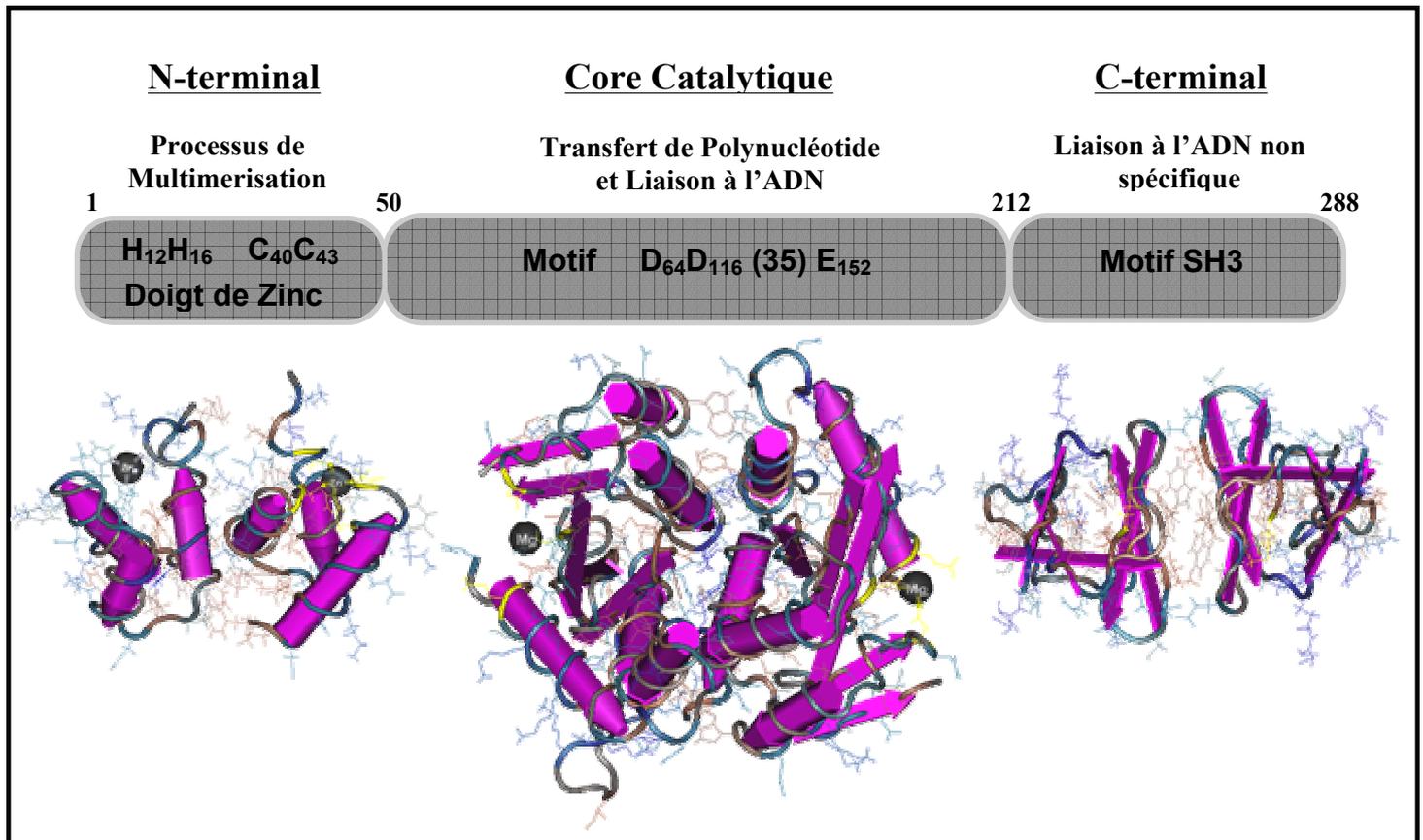


Figure 27 : Structure et fonction des 3 domaines de l'IN

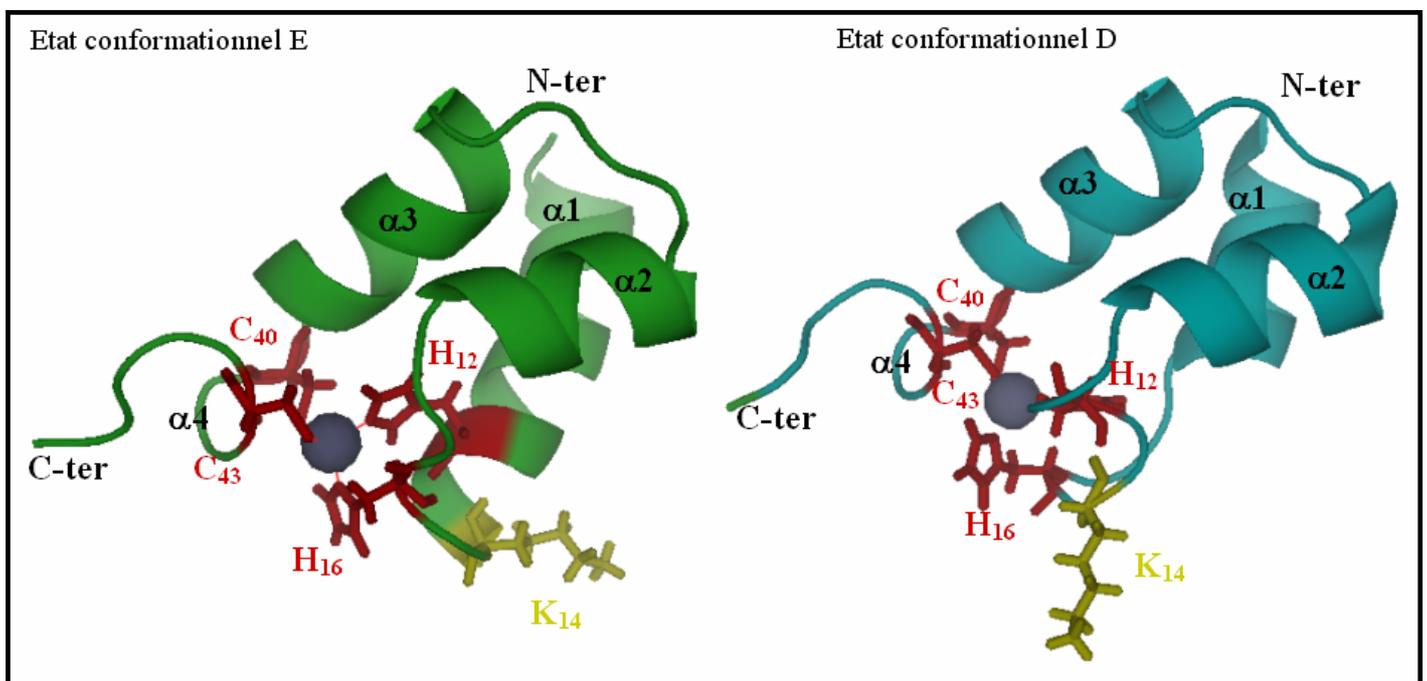


Figure 28 : Structures du domaine N-ter de l'IN du VIH-1

Etat conformationnel E (PDB : 1WJC), vert. Etat conformationnel D (PDB : 1WJA), bleu. En rouge, les résidus coordonnateurs du  $Zn^{2+}$  : His<sub>12</sub>-His<sub>16</sub>-Cys<sub>40</sub>-Cys<sub>43</sub>. En jaune, la Lys 14 qui interagit avec l'ADN. Entre les 2 formes, changement du positionnement de la chaîne latérale de la Lys14 et de la longueur de l'hélice  $\alpha$  1. D'après (Cai et al., 1997)

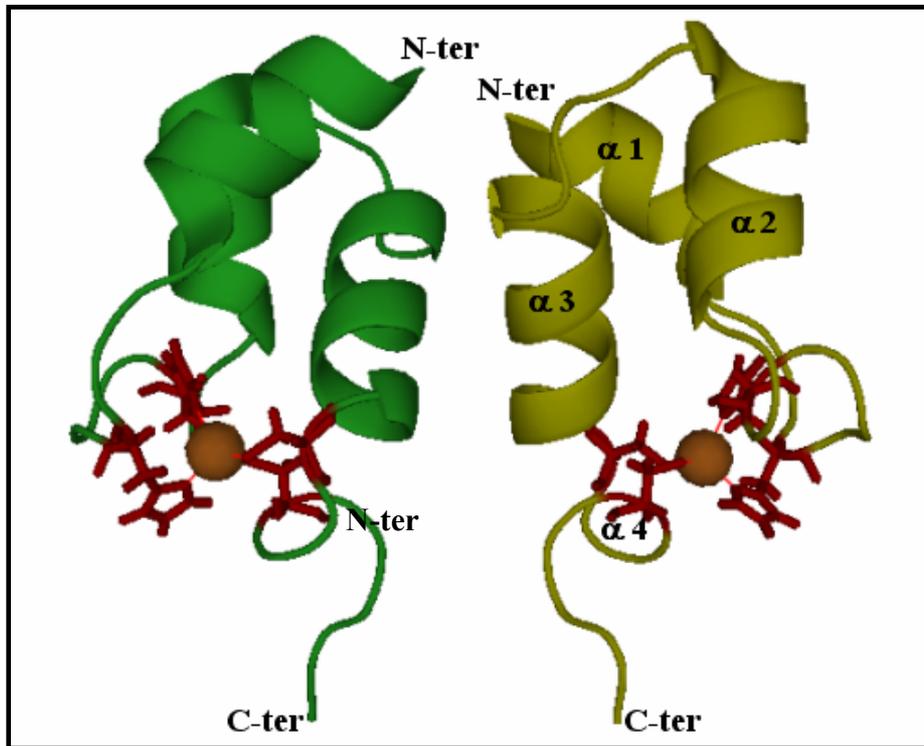
## II.B. Structures et fonctions des trois domaines de l'intégrase

Dans un premier temps, les structures des trois domaines ont été obtenues séparément entre 1994 et 1997 que se soit par diffraction aux rayons X ou par RMN (core catalytique (Dyda *et al.*, 1994), domaine C-Ter : (Lodi *et al.*, 1995), (Eijkelenboom *et al.*, 1995), domaine N-Ter : (Cai *et al.*, 1997)) (cf. figure 27). C'est en 2000 et 2001 que deux équipes arrivent à cristalliser les doubles domaines N-ter / CC (Wang *et al.*, 2001) et CC / C-Ter (Chen *et al.*, 2000a). Cependant, l'agencement structural des trois domaines reste inconnu car les propriétés biochimiques et agrégatives de l'IN entière, même en présence de mutations ponctuelles ou de produits additionnels favorisant sa solubilité, empêchent l'obtention de cristaux de bonne qualité avec un bon pouvoir de diffraction, et donc de la structure à haute résolution résultante.

### II.B.1. Le domaine N-Terminal

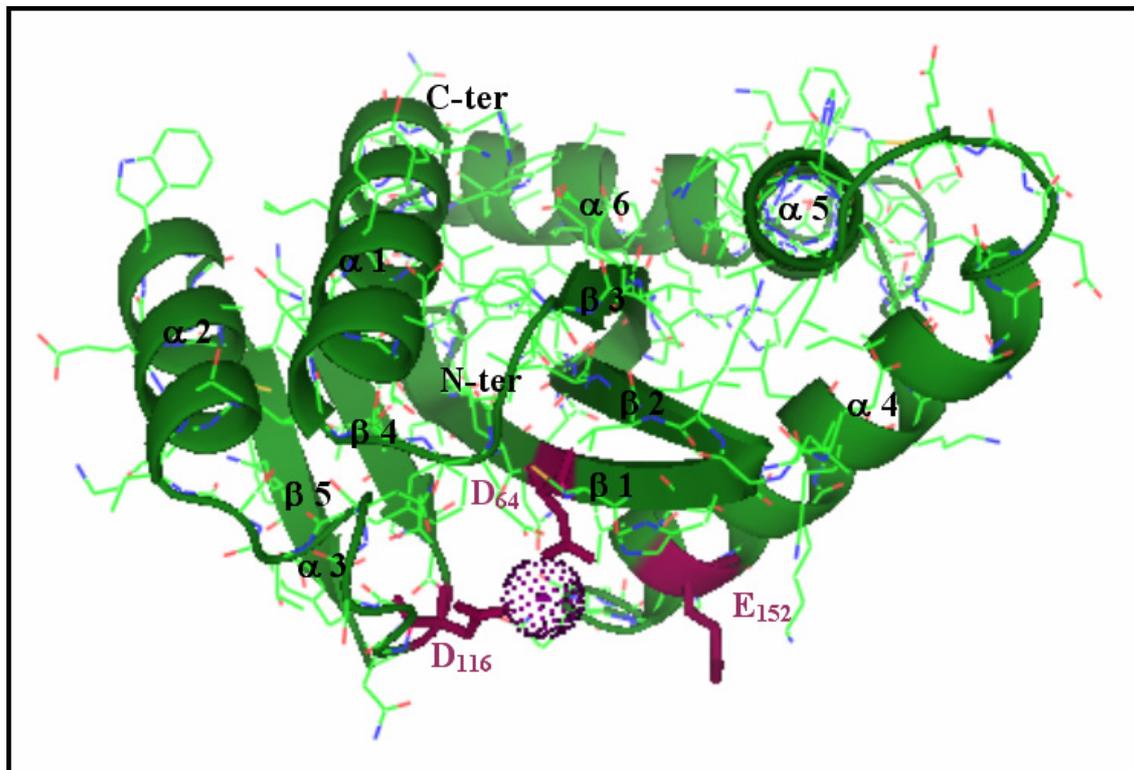
Le domaine N-ter est formé par les 49 premiers résidus de l'IN. Ce domaine n'est structuré qu'en présence de zinc. Lorsqu'il est structuré, il comprend quatre hélices  $\alpha$  ( $\alpha$  1 : résidus 2 - 14,  $\alpha$  2 : résidus 19 - 25,  $\alpha$  3 : résidus 30 - 39 et  $\alpha$  4 : résidus 41 - 45). C'est en 1997 que sa structure a été obtenue par RMN par Cai *et al.* ((Cai *et al.*, 1997), code PDB : 1WJA) (cf. figure 28 et 29). Ce domaine forme un pseudo-doigt de zinc de type HHCC composé des résidus H<sub>12</sub>-H<sub>16</sub>-C<sub>40</sub>-C<sub>43</sub> (motif His-X<sub>aa3</sub>-His-X<sub>aa20/30</sub>-Cys-X<sub>aa2</sub>-Cys), très conservés chez les intégrases rétrovirales et les rétrotransposons eucaryotes. Ce motif, dont la structure secondaire est atypique pour un doigt de zinc, normalement constitué de deux hélices et d'un coude, permet la fixation d'un cation divalent Zn<sup>2+</sup> avec une stoechiométrie d'un ion par monomère (cf. figure 28). L'affinité de l'apoprotéine pour le cation a été déterminée par dialyse à l'équilibre :  $K_{d,apparent} \approx 6\mu\text{M}$  (pH = 7,5, NaCl = 1M) (Zheng *et al.*, 1996). L'ion divalent a pour fonction de stabiliser l'ensemble hélice-boucle-hélice et donc l'arrangement tridimensionnel du domaine.

Il est intéressant de noter que le complexe N-ter / Zn<sup>2+</sup> existe sous deux états conformationnels nommés E et D (cf. figure 28). Les deux états se différencient par la coordination du cation par les deux histidines 12 et 16. Dans la forme E, prédominante à 300K et à pH 7,4, l'His16 coordine le zinc via son atome N $\delta$ 1. L'His 12 est alors enfouie dans la structure et coordine le zinc par son N $\epsilon$ 2 et forme en plus une liaison hydrogène par



**Figure 29 : Structure du dimère du domaine N-ter (forme D)**

Structure d'après (Cai et al., 1997), en rouge, les résidus coordonnant le  $Zn^{2+}$  (H<sub>12</sub>-H<sub>16</sub>-C<sub>40</sub>-C<sub>43</sub>) et en orange, les atomes de zinc. (PDB : 1WJA)



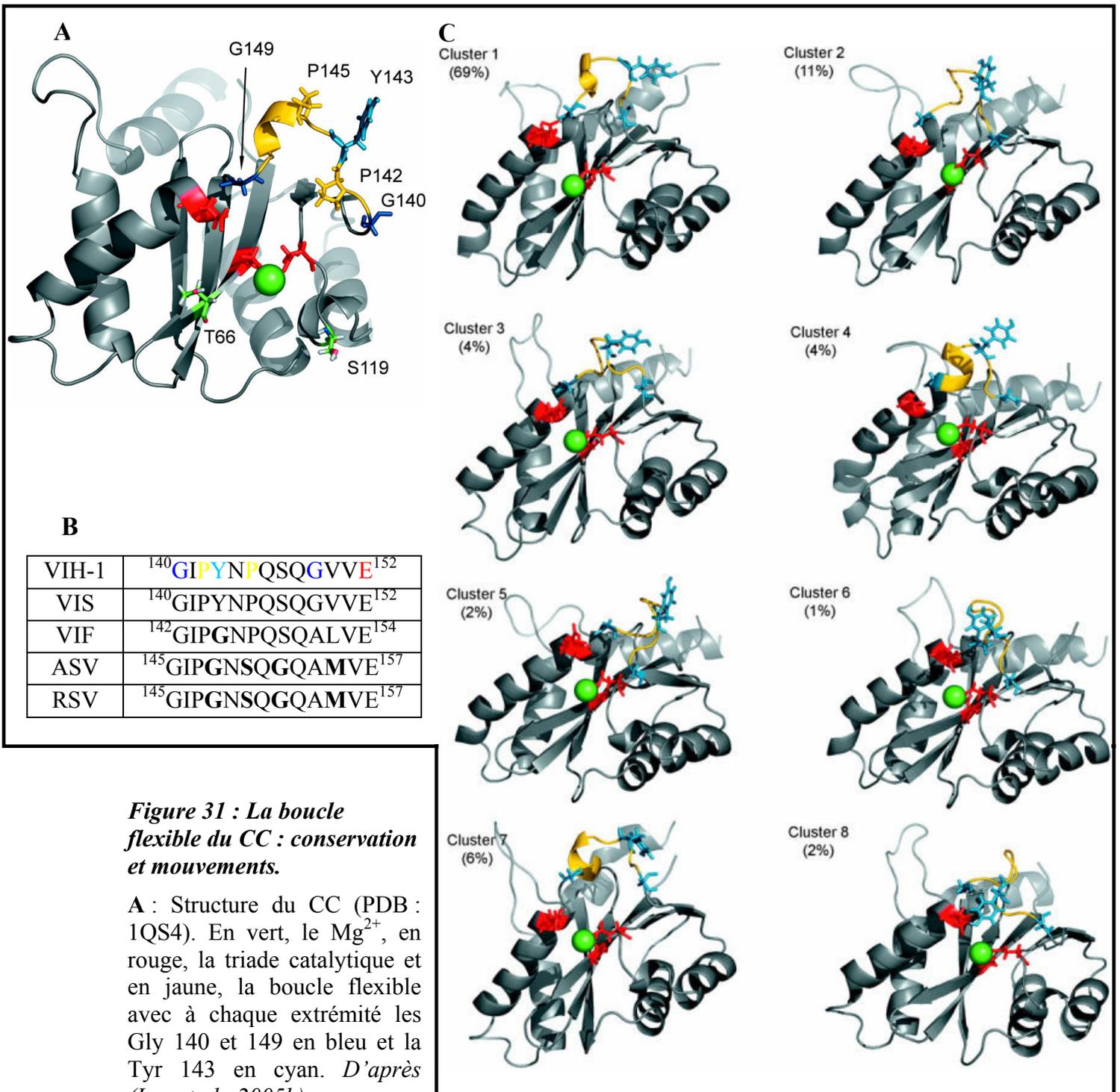
**Figure 30 : Structure du monomère du CC de l'IN du VIH-1**

En magenta, les trois résidus de la triade catalytique : D64, D116 et E152. Coordination du  $Mg^{2+}$  par les deux Asp 64 et 116. (PDB : 1BL3). D'après (Maignan et al., 1998)

l'intermédiaire de son proton N $\delta$ 1H avec l'atome S de la Met 22. Dans cette conformation, l'His 16 est exposée au solvant. Dans la forme D, les positions relatives des deux histidines sont inversées. De ce fait, l'His 12 se retrouve exposée au solvant et coordine le zinc par son N $\delta$ 1. Les deux formes se différencient également par des changements de conformation plus large dans l'environnement de ces deux histidines notamment au niveau du polypeptide allant des résidus 9 à 18. En effet, l'hélice  $\alpha$ 1 qui va des résidus 2 à 14 dans la forme E est scindée en une hélice  $\alpha$  qui va des résidus 2 à 8 suivit d'un tour  $\alpha$  hélicoïdal qui comprend les résidus 14 à 17 dans la forme D.

Le domaine N-ter dimérise en solution. L'interface entre les deux monomères est identique pour les deux formes. Elle est surtout constituée par des interactions hydrophobes et fait intervenir la fin de l'hélice  $\alpha$ 1 et les hélices  $\alpha$ 3 et  $\alpha$ 4 (cf. figure 29). Le domaine N-ter interviendrait dans la multimérisation de l'IN. Des expériences ont montré que le zinc dans une stœchiométrie 1 : 1 stimule l'activité de 3'-processing et de transfert de brin de l'IN en présence de Mg<sup>2+</sup> alors qu'en Mn<sup>2+</sup>, ces activités ne semblent pas pâtir de l'absence du cation. En désintégration, le zinc ne semble pas non plus jouer de rôle primordial dans l'efficacité de la réaction ((Zheng *et al.*, 1996); (Lee *et al.*, 1997)). Le zinc pourrait contribuer, de par son effet structurant sur le domaine N-ter, à la reconnaissance spécifique des extrémités de l'ADN. De ce fait, dans le contexte de l'intégrase entière, le domaine N-ter pourrait jouer un rôle dans les interactions IN / ADN. Des expériences de foot-printing sur protéine permettant de différencier chacun des monomères ont permis de mettre en évidence une interaction entre la Lys14 et l'ADN<sub>viral</sub>. En effet, lorsque le monomère est chimiquement lié à l'ADN<sub>viral</sub>, cette lysine présente un changement radical de sensibilité à la digestion trypsique témoin d'une protection de cette lysine par l'ADN (Zhao *et al.*, 2008). Ceci montre donc que le domaine N-ter pourrait non seulement intervenir dans la multimérisation de l'intégrase entière, mais également dans la reconnaissance de l'ADN<sub>viral</sub>.

Ce domaine pourrait être en partie responsable des propriétés agrégatives observées lors des purifications de l'IN du VIH-1 (Ellison *et al.*, 1995). En effet, lors de la construction de chimères entre l'IN du VIH-1 et celle de PFV-1 (Primate Foamy Virus de type 1), les deux constructions N-Ter<sub>VIH</sub> / CC+C-Ter<sub>PFV</sub> et N-Ter<sub>PFV</sub> / CC+C-Ter<sub>VIH</sub> possèdent des propriétés de solubilités proches des intégrases qui ont donné le domaine N-Ter (Lee *et al.*, 2005a). Dans un contexte cellulaire, le domaine N-ter est la région de l'IN qui possède la plus forte affinité pour le facteur de transcription INI1 qui est un composant du CPI. Ce domaine pourrait donc



**Figure 31 : La boucle flexible du CC : conservation et mouvements.**

**A** : Structure du CC (PDB : 1QS4). En vert, le  $Mg^{2+}$ , en rouge, la triade catalytique et en jaune, la boucle flexible avec à chaque extrémité les Gly 140 et 149 en bleu et la Tyr 143 en cyan. *D'après (Lee et al., 2005b).*

**B** : Alignement des structures de la boucle flexible de plusieurs IN rétrovirales. En gras, les résidus différents de ceux observés chez l'IN du VIH-1. *D'après (Johnson et al., 2006).*

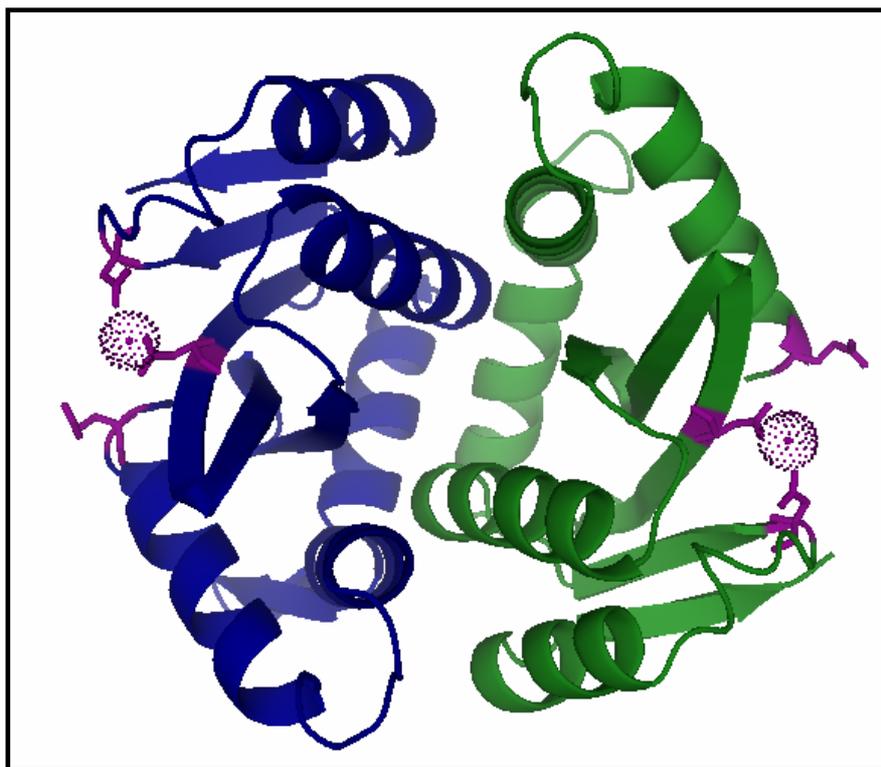
**C** : Modélisation des mouvements de la boucle flexible. Chaque cluster est représenté ainsi que sa probabilité de positionnement. Sur ces clusters, on remarque les changements de position de la Tyr 143 en cyan. *D'après (Lee et al., 2005b).*

avoir une importance dans les interactions protéines / protéines qui interviennent dans la formation du CPI. Il stabilise également l'interaction qui existe entre le CC de l'IN du VIH-1 et LEDGF/p75 (Maertens *et al.*, 2003).

### ***II.B.2. Le Core Catalytique***

Ce domaine central, résistant à la protéolyse ménagée, formé par les résidus allant de 50 à 212, contient un motif DD-35-E très conservé chez les intégrases rétrovirales ainsi que chez les transposases trouvées chez les organismes procaryotes. Il comprend les résidus D<sub>64</sub>-D<sub>116</sub>-E<sub>152</sub> (Engelman *et al.*, 1992). Cette triade catalytique fait partie d'une poche hydrophobe qui fixe un cation divalent Mn<sup>2+</sup> ou Mg<sup>2+</sup> qui est nécessaire à l'activité enzymatique. Ce motif est responsable des activités catalytiques de l'IN, toute mutation de l'un de ces trois résidus entraîne l'abolition de toute activité enzymatique de l'IN (3'-processing, transfert de brin, intégration concertée, coupure de type palindromique et désintégration). Sa structure comprend un feuillet central de cinq brins β flanqué au niveau de chacune de ses deux faces de trois hélices α (cf. figure 30). Cette organisation présente beaucoup de similarité avec la structure observée pour la RNase H, ou pour certains domaines de type adénosine triphosphatase de plusieurs protéines comme l'hexokinase ou la protéine chaperonne de 70 kDa (HSP 70).

Lors de la première structure cristalline obtenue par Dyda *et al.* en 1994 (Dyda *et al.*, 1994), trois séries de résidus étaient non résolues : les neuf premiers et les quatre derniers résidus de la séquence ainsi qu'une zone comprise entre les résidus 141 et 153. Ces résidus comprennent non seulement le troisième résidu de la triade catalytique, mais également une boucle flexible importante pour l'activité catalytique de l'IN (cf. figure 31 A) dont la séquence est conservée chez les intégrases rétrovirales (cf. figure 31 B). Son importance a été notamment montrée par des mutations ponctuelles au niveau des deux glycines qui la délimite (Gly 140 et Gly 149), et l'activité de désintégration du double mutant résultante est diminuée d'environ 90% (Greenwald *et al.*, 1999). Cette boucle possède une géométrie stabilisée par deux résidus prolines (142 et 145) situés autour de la zone responsable en grande partie du mouvement d'ouverture ou de fermeture de la boucle tel un couvercle sur le site actif. Ces résidus adoptent une conformation de type boucle β-α en position dite ouverte qui s'inverse lors du mouvement de fermeture (Lee *et al.*, 2005b) (cf. figure 31 C). Ce mouvement permet notamment l'orientation vers le site catalytique du résidu Tyr 143 qui permettrait de stabiliser



**Figure 32 : Structure du dimère du CC, en présence de  $Mg^{2+}$**

En magenta sur chacun des deux monomères, la triade catalytique avec l'atome de  $Mg^{2+}$  coordonné par les Asp 64 et 116. (PDB : 1BL3). D'après (Maignan et al., 1998)

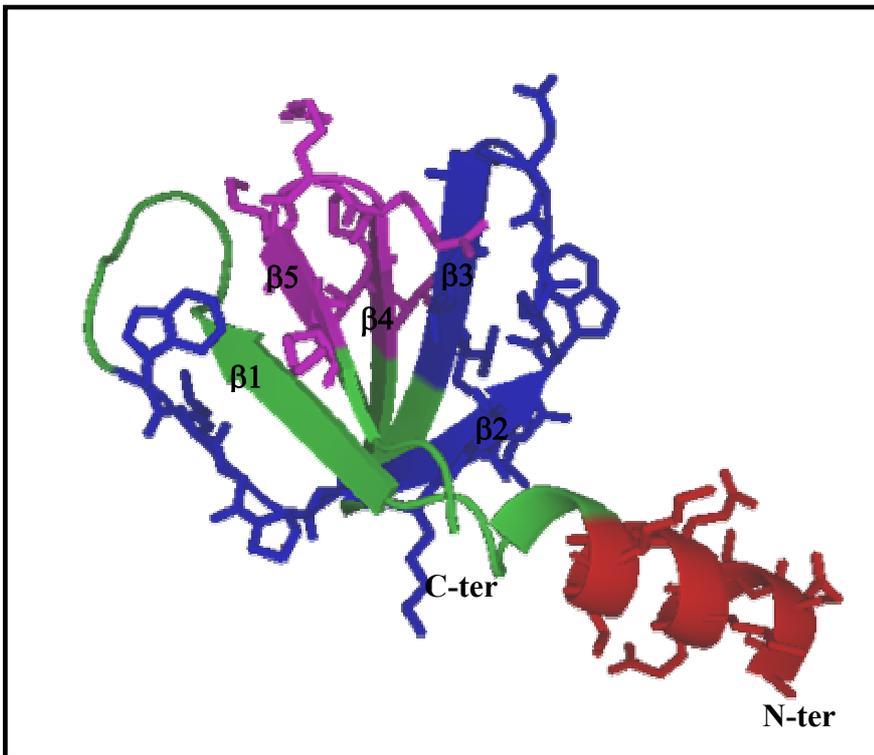
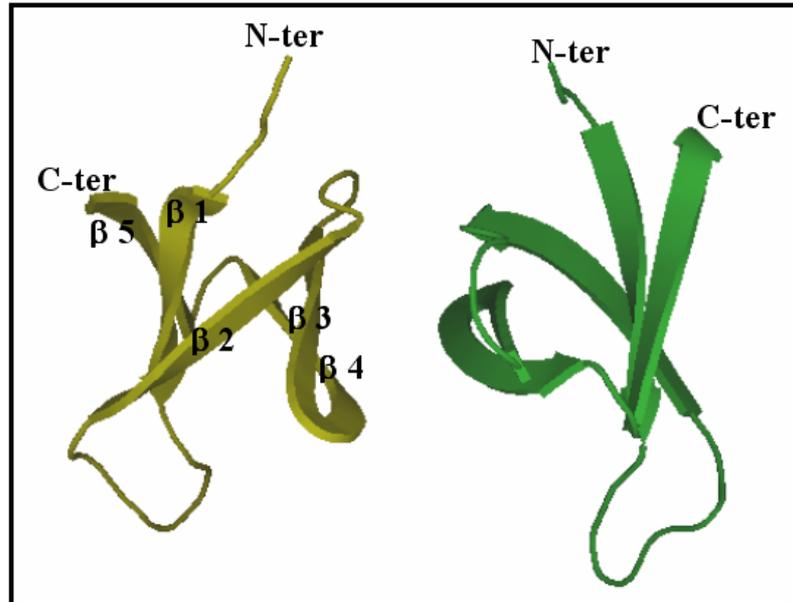
la molécule d'eau activée. Des études de mutagenèse dirigée de ce résidu ont montré que l'attaque nucléophile réalisée par la molécule d'eau en présence de la Tyr 143, pouvait être alors réalisée par des molécules de taille plus importante comportant un groupement -OH comme le glycérol ((De Luca *et al.*, 2005); (van Gent *et al.*, 1993)). Un deuxième résidu de cette boucle est primordial pour l'activité catalytique de l'IN : la Gln 148. Plusieurs expériences (cross-link, mutagenèse dirigée,...) montrent clairement que ce résidu interagit de manière spécifique avec l'ADN<sub>viral</sub>. Il formerait une liaison hydrogène avec l'azote 3 de la cytosine présente à l'extrémité 5' de l'ADN<sub>viral</sub>. Lorsque la coupure endonucléolytique a eu lieu, cette cytosine peut alors réaliser une rotation par l'intermédiaire du squelette sucre/phosphate et former cette liaison hydrogène. Cette rotation couplée à un mouvement de la boucle flexible, provoquerait un changement conformationnel de l'ADN qui permettrait de rendre plus accessible l'extrémité clivée 3'-OH de l'ADN<sub>viral</sub> pour la réaction suivante de transfert de brin (Johnson *et al.*, 2006).

Différentes structures du CC ont été obtenues. Elles ont notamment permis de tracer la conformation de la boucle flexible, de positionner la chaîne latérale du résidu Glu 152 et de confirmer le positionnement du cation divalent ( $Mg^{2+}$  ((Goldgur *et al.*, 1998), (Maignan *et al.*, 1998))). Ce dernier établit des liaisons de coordinations avec d'une part les oxygènes des deux Asp (64 et 116) de la triade catalytique et d'autre part avec quatre molécules d'eau. Ces quatre liaisons de coordinations forment le plan carré de l'octaèdre de coordination du cation divalent. La résolution dans l'axe perpendiculaire ne permet pas de donner plus de détails, mais met en évidence une certaine mobilité du cation. Le troisième résidu de la triade catalytique (Glu 152) interagirait avec la guanine du CAGT en 3' de l'ADN<sub>viral</sub>, qui suit directement la liaison phosphodiester clivée.

Les structures cristallines montrent que le CC est sous forme de dimère comme cela a été observé pour le domaine N-ter en solution (cf. figure 32). La zone de contact est formée d'une large interface excluant toute molécule de solvant de  $1300 \text{ \AA}^2$  par sous-unités. Cette interface comporte un nombre conséquent de pont salins et de liaisons hydrogènes formés par des résidus appartenant au feuillet  $\beta 3$  et aux hélices  $\alpha 1$ ,  $\alpha 3$ ,  $\alpha 5$  et  $\alpha 6$ . Le CC est impliqué dans les reconnaissances respectivement spécifique des extrémités de l'ADN<sub>viral</sub> et non spécifique de l'ADN hôte ((Engelman *et al.*, 1994) , (Ellison *et al.*, 1995), (Chen *et al.*, 2006)). Cependant, l'espace entre les sites actifs dans le dimère n'est pas compatible avec l'espace nécessaire à l'insertion de l'ADN<sub>viral</sub>. En effet, les sites d'insertion sur chaque brin

**Figure 33 : Structure du dimère du domaine C-ter**

Structure résolue du domaine C-ter allant des résidus 222 à 272. (PDB : 1QMC). D'après (Eijkelenboom et al., 1995)



**Figure 34 : Régionalisation du domaine C-ter de l'IN du VIH-1**

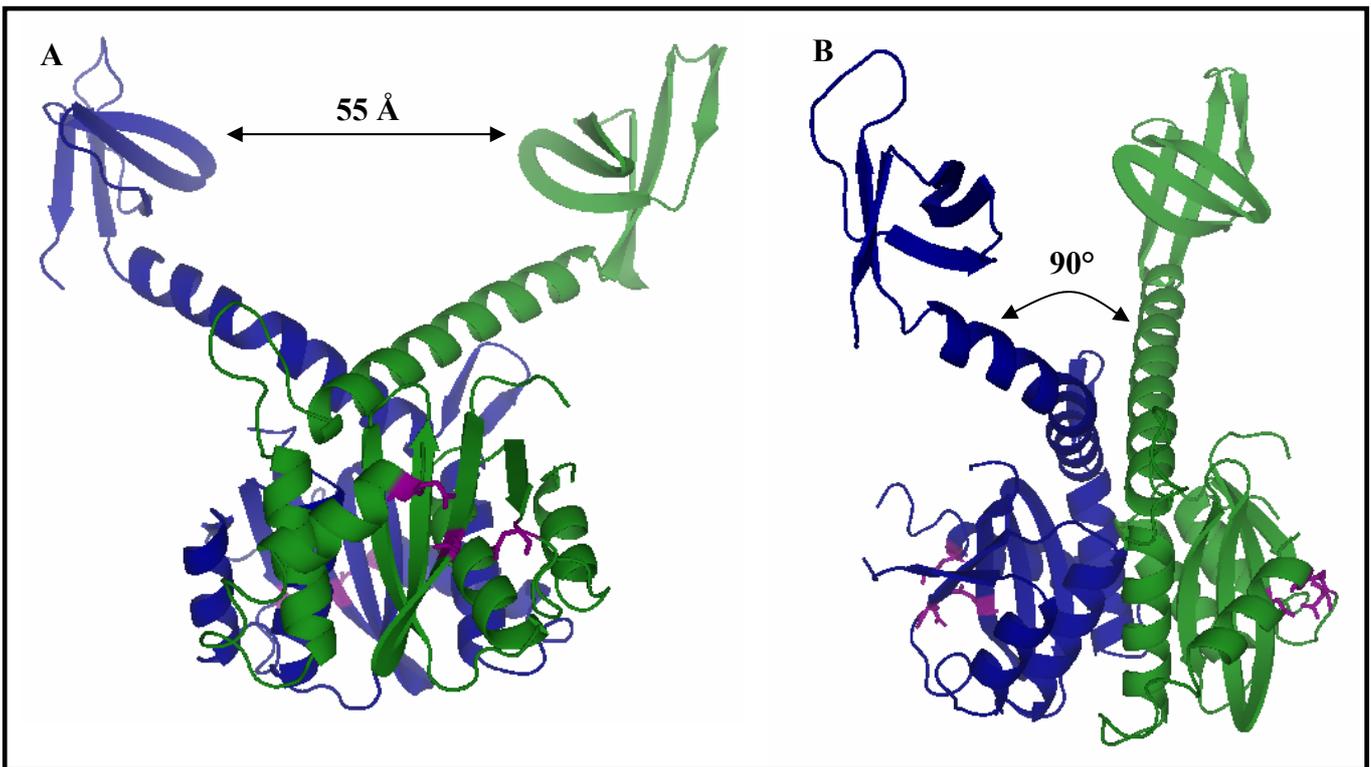
Structure des résidus 212 à 272, issue de la structure du double domaine CC / C-ter. En rouge, les résidus de la région Q, en bleu, la région C et en violet, la région N. Structuration de la partie N-terminale du domaine en hélice  $\alpha$ . (PDB : 1EX4). D'après (Chen et al., 2000a)

d'ADN cellulaire sont séparés de 5 paires de bases soit environ 15 Å, alors que l'espace entre les sites actifs dans les structures cristallines est de plus de 30 Å (Dyda *et al.*, 1994). Ceci suggère donc un niveau supérieur de multimérisation de l'IN lors de l'étape de l'intégration, qui pourrait nécessiter la formation d'un tétramère.

### ***II.B.3. Le domaine C-Terminal***

Le domaine C-ter comprend les 76 derniers résidus. Ce domaine, qui dimérise lui aussi en solution, est le moins conservé des trois domaines au sein de la famille des intégrases rétrovirales. Cependant au niveau structural, le domaine se compose toujours de 5 brins  $\beta$  ( $\beta_1$  : résidus 223 à 228,  $\beta_2$  : résidus 235 à 245,  $\beta_3$  : résidus 248 à 252,  $\beta_4$  : résidus 256 à 260,  $\beta_5$  : résidus 265 à 268) qui forment deux feuillets  $\beta$  antiparallèles ( $\beta_1$ - $\beta_2$ - $\beta_5$  et  $\beta_3$ - $\beta_4$ ) (cf. figure 33). L'organisation tridimensionnelle de ce domaine est similaire à celle des domaines de type SH3, qui sont généralement retrouvés dans de nombreuses protéines cellulaires impliquées dans la transduction de signaux de croissance, de migration ou de différenciation cellulaire ((Eijkelenboom *et al.*, 1995); (Lodi *et al.*, 1995)).

Il est établi qu'il contient des déterminants pour la multimérisation et qu'il fixe l'ADN viral de manière non spécifique ((Eijkelenboom *et al.*, 1995); (Lutzke and Plasterk, 1998); (Engelman *et al.*, 1994); (Jenkins *et al.*, 1996)). Ce domaine serait porteur d'un signal d'adressage au noyau atypique qui jouerait un rôle dans l'import nucléaire du CPI ((Gallay *et al.*, 1997); (Bouyac-Bertoia *et al.*, 2001)). Cependant, les études sur l'implication de l'IN dans l'import nucléaire restent assez controversées ((Dvorin *et al.*, 2002); (Limon *et al.*, 2002); (Devroe *et al.*, 2003)). Le domaine C-Ter contient également trois régions conservées chez les intégrases rétrovirales nécessaires à la réplication virale, nommées C (<sub>235</sub>WKGPALLWKGEAVV), N (<sub>259</sub>VVPRRKAK) et Q (<sub>211</sub>KELQKQITK) (Cannon *et al.*, 1996) (cf. figure 34). Des expérimentations de mutagenèse dirigée couplées à des infections de cellules par des virions pseudo-typés montrent que les tri-lysines des régions conservées C et Q seraient importantes de manière directe ou indirecte pour la multiplication virale à deux niveaux clef des étapes précoces que sont la reverse transcription et l'import nucléaire (Ao *et al.*, 2005).



**Figure 35 : Structure du double domaine CC/C-ter de l'IN**

En vert, le monomère A, en bleu, le B et en magenta, la triade catalytique.

**A** : Vue du dimère de face. Distance de 55 Å entre les deux C-ter.

**B** : Vue du dimère de profil. Angle de 90° entre les deux hélices  $\alpha$  6 à partir de la Thr 210 du monomère B. (PDB : 1EX4). D'après (Chen et al., 2000a).

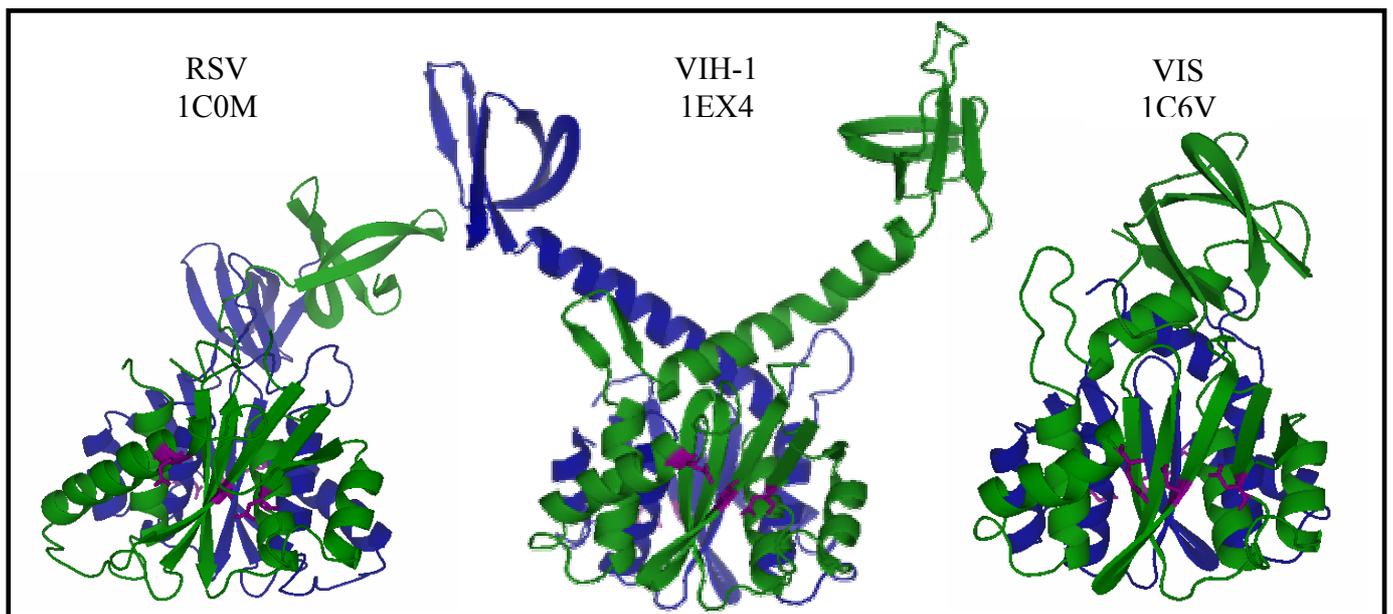
## II.C. Structure de l'intégrase : des doubles domaines à la modélisation

### II.C.1. Structure des doubles domaines

#### II.C.1.a. Structure du double domaine core catalytique / C-terminal

Les propriétés agrégatives de l'IN du VIH-1 entière ou tronquée ont limité son étude structurale aux domaines isolés. L'introduction de plusieurs mutations ponctuelles dans sa séquence afin d'améliorer sa solubilité a permis dans un premier temps l'obtention de cristaux des domaines isolés, puis celles des constructions des doubles domaines de l'IN. La mutation F185K est connue pour améliorer la solubilité de l'IN (Jenkins *et al.*, 1995). Elle a permis notamment d'augmenter la solubilité du CC et de le cristalliser. Pour améliorer la solubilité, les mutations ponctuelles W131D et F139D permettent d'éliminer de la surface du CC deux résidus hydrophobes, et donc de diminuer également les propriétés agrégatives de l'IN par diminution des interactions hydrophobes protéines – protéines responsables du phénomène d'agrégation. Enfin les deux mutations C56S et surtout C280S permettent de limiter la formation de ponts disulfure intermoléculaires. La solubilité de l'IN pentamutée a notamment permis d'obtenir des cristaux de l'intégrase entière, mais seulement avec une résolution de la structure à 8 Å (Chen *et al.*, 2000a). A partir de la séquence de cette IN pentamutée, le double domaine, CC / C-ter (IN<sup>52-288</sup>) a pu être également purifié dans des concentrations compatibles avec la cristallisation. Les cristaux obtenus ont permis de résoudre la structure du double domaine avec une résolution de 2,8 Å. La structure du double domaine se compose d'un dimère. Les interactions intermoléculaires du dimère se font uniquement au niveau du core catalytique, dont les interactions et la structure sont quasiment identiques à la structure obtenue pour le domaine central isolé avec pour les C  $\alpha$  une déviation d'environ 0,6 Å entre les monomères. Les domaines C-terminaux se trouvent positionnés de manière symétrique par rapport à un axe de symétrie plan. Si la structure tertiaire du domaine C-ter est identique à celle déjà obtenue pour ce domaine seul en solution, sa structure quaternaire est différente puisque ce domaine n'est plus sous forme dimérique.

Le dimère du double domaine CC / C-ter présente donc une forme en Y avec un écart entre les deux domaines C-ter de 55 Å et un angle d'environ 90° entre les deux hélices  $\alpha$  6 (résidus : 195 à 220) qui relie chacun des domaines à son CC (cf. figure 35). C'est au niveau

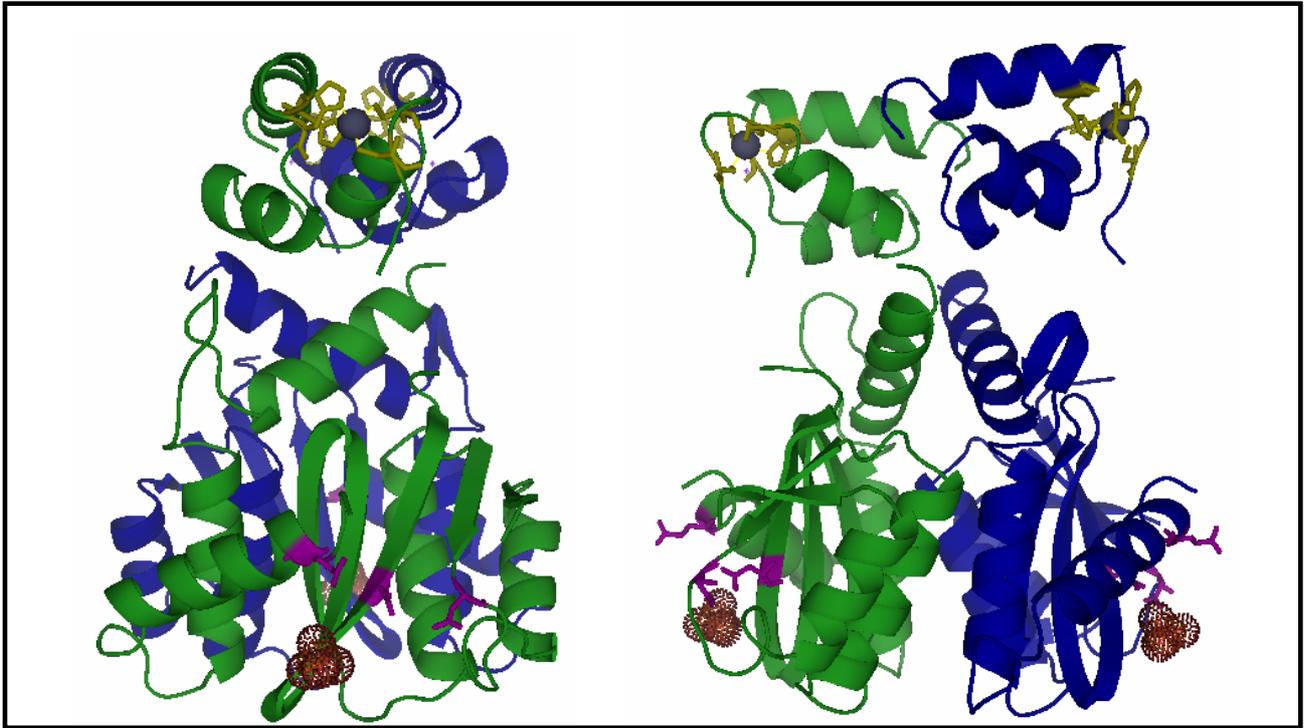


**Figure 36 : Comparaison des doubles domaine CC/C-ter des IN du VIH-1, RSV et VIS.**

En vert le monomère A, en bleu le B et en magenta la triade catalytique. D'après (Chen et al., 2000a).

de la Thr 210 que l'angle de 90° par rapport à l'hélice du monomère A est initié au niveau de l'hélice  $\alpha$  6 du monomère B. Au niveau de la résolution de la structure, trois zones présentent une flexibilité trop importante pour obtenir une résolution interprétable : des résidus 138 à 149 qui correspondent à la boucle flexible du site actif, des résidus 186 à 194 qui se situent entre les hélices  $\alpha$  5 et  $\alpha$  6 et des résidus 271 à 288. La flexibilité de la zone allant des résidus 186 à 194 pourrait refléter un phénomène conformationnel clef dans la dynamique fonctionnelle de l'activité catalytique séquentielle de l'IN. Cette liaison entre le domaine C-ter et le CC montre une grande variabilité notamment lorsque l'on compare cette structure avec les structures résolues des autres doubles domaines d'intégrases rétrovirales cristallisées (RSV<sup>49-286</sup> et SIV<sup>50-293</sup>) (cf. figure 36). Dans le cas de l'IN de RSV (Yang *et al.*, 2000), vont dans la même direction, ce qui aboutit à un positionnement des deux domaines C-ter du même côté de l'axe de symétrie déterminé par le dimère de CC. Ce positionnement permet de stabiliser la structure en masquant un nucleus hydrophobe comprenant les résidus Trp 213 et Ile 209. L'absence du domaine N-ter, qui pourrait d'une part masquer ce core hydrophobe et d'autre part jouer un rôle stabilisateur de la position du C-ter, peut expliquer ce positionnement asymétrique potentiellement artefactuel des deux C-ter. Pour l'IN du SIV (Chen *et al.*, 2000b), seul un des quatre C-ter de la maille cristallographique est résolu. Cette faible résolution montre surtout une grande flexibilité quand au positionnement du domaine C-ter par rapport au dimère de CC. Ces observations comparatives de ces trois structures de doubles domaines montrent une certaine flexibilité, qui pourrait être symptomatique d'un changement de conformation des hélices  $\alpha$  6 lors de la fixation sur l'ADN ou de l'une activité catalytique de l'IN.

Dans la structure obtenue avec le double domaine C-ter / CC de l'IN de VIH-1, la majorité des contacts entre les différents dimères se font au niveau des domaines C-ter par quatre types de contacts : dans les deux premiers cas, les interactions se font entre les brins  $\beta$ 2- $\beta$ 3- $\beta$ 4 positionnés parallèlement avec une molécule de CHAPS à l'interface pour stabiliser les interactions qui comprennent en plus des interactions hydrophobes des liaisons hydrogènes. Il en est de même pour le troisième type d'interaction qui comprend les feuillets  $\beta$ 1- $\beta$ 2- $\beta$ 5 de manière antiparallèle avec aussi une molécule de CHAPS stabilisatrice. Le dernier type d'interaction comprend les feuillets  $\beta$ 2- $\beta$ 3 d'un C-ter et le feuillet  $\beta$ 5 de l'autre, avec un positionnement des dimères d'IN cote à cote (Chen *et al.*, 2000a).



**Figure 37 : Structure du double domaine N-ter/CC de l'IN**

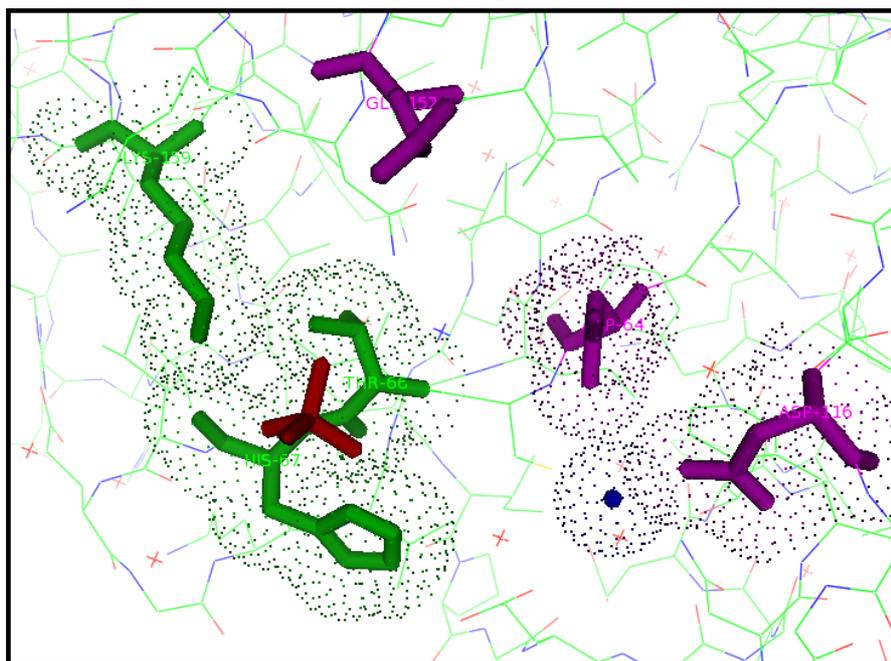
Vue de face et de profil (rotation de 90°)

En vert, le monomère A, en bleu, le monomère B, en gris, les atomes de zinc et en jaune, les His et Cys qui les coordinent, en magenta, la triade catalytique et en orange, les phosphates. (PDB : 1K6Y). D'après (Wang *et al.*, 2001).

### II.C.1.b. Structure du double domaine N-Ter / Core catalytique

L'IN<sup>1-212</sup> cristallisée comporte trois mutations (W131D, F139D et F185K) qui ont permis d'améliorer sa solubilité et sa propension à cristalliser et ce sans modifier l'activité 3'-processing de l'intégrase entière triple-mutante ainsi que l'activité de désintégration de la protéine tronquée (Wang *et al.*, 2001). Les structures tertiaires du domaine N-ter et du CC monomériques en solution restent très similaires de celles observées dans le cadre du double domaine (cf. figure 37).

Cependant, l'interface entre les deux domaines N-ter du dimère est différente de l'interface d'interaction observée pour le domaine N-ter seul en solution. L'interface au niveau du dimère de double domaine est plus restreinte (quelques résidus de la première et de la troisième hélice  $\alpha$ ) et plus hydrophobe que celle du dimère du domaine N-ter seul en solution. Seuls les résidus Phe 1, Leu 2, Pro 29, Val 31 et Val 32 sont communs aux deux interfaces. Pour passer de la forme dimère N-ter seul à dimère N-ter / CC, il faut réaliser sur un des deux monomères une rotation de 180° centrée sur l'interface hydrophobe suivi d'une translation. Au niveau du dimère du double domaine, les résidus impliqués comprennent en plus Ile 5, qui augmente un peu plus l'hydrophobicité de cette interface. A noter que dans la structure du double domaine, les résidus 47 à 55 qui assurent la connexion entre le domaine N-ter et le CC ne présentent aucune structuration stable dans aucun des quatre monomères observés dans la maille cristallographique. Ceci dit, la distance observée entre le CC et le domaine N-ter (9 ou 13 Å selon la sous-unité observée) est compatible avec la présence d'une liaison comportant neuf résidus. Cette dernière rend d'ailleurs impossible la dimérisation du N-ter comme celle observée pour le domaine N-ter seul en solution puisqu'il faudrait dans ce cas là une distance de 35 Å environ, qui est incompatible avec le nombre de 9 résidus non résolus. Contrairement à l'interface présente entre les deux domaines N-ter, l'interface entre le domaine N-ter / CC est beaucoup plus hydrophile. Elle comprend notamment des interactions entre les chaînes latérales des résidus Arg 20, Lys 34, Gln 209, Thr 206 et Glu 212. Cette surface est d'environ 900 Å<sup>2</sup>. Les dimères ne sont pas positionnés de manière identique. En effet, l'orientation des deux domaines N-ter diffère d'environ 15°. Les interactions entre le domaine N-ter et le CC présentent une asymétrie entre les deux sous-unités d'un dimère. Cependant, comme cette asymétrie est également visible sur les deux autres sous-unités, l'arrangement asymétrique ne semble pas être un artéfact du à la formation du cristal, mais pourrait bien avoir une relevance biologique.



**Figure 38 : Positionnement du phosphate au niveau du site actif**

En rouge, le phosphate, en vert, les trois résidus coordinateurs (His 67, Thr 66 et Lys 159), en bleu la molécule d'eau, coordonnée par les deux Asp 64 et 116 de la triade catalytique (en magenta), qui pourrait être positionnée comme le cation divalent. *D'après (Wang et al., 2001).*

La structure cristalline révèle clairement la présence d'un ion phosphate pour chacun des quatre monomères dans la région du site actif (à 7 Å) qui mimerait une partie du squelette phosphate de l'ADN lié. Chacun des quatre phosphates est coordonné par 5 liaisons avec la chaîne latérale des résidus Thr 66, His 67, Lys 159, la chaîne principale de His 67 et pour finir une molécule d'eau qui occupe très probablement la place d'un cation métallique divalent (cf. figure 37 et 38). De plus, la liaison du phosphate entraînerait un changement conformationnel du site actif, notamment de la position dans l'espace de la triade catalytique D, D-35-E. Cette hypothèse est renforcée par la conformation de la boucle contenant Asp 116 qui est identique à celle observée pour la structure du dimère du CC seul en présence de  $Mg^{2+}$ .

### ***II.C.2. Modélisation de la structure de l'intégrase entière***

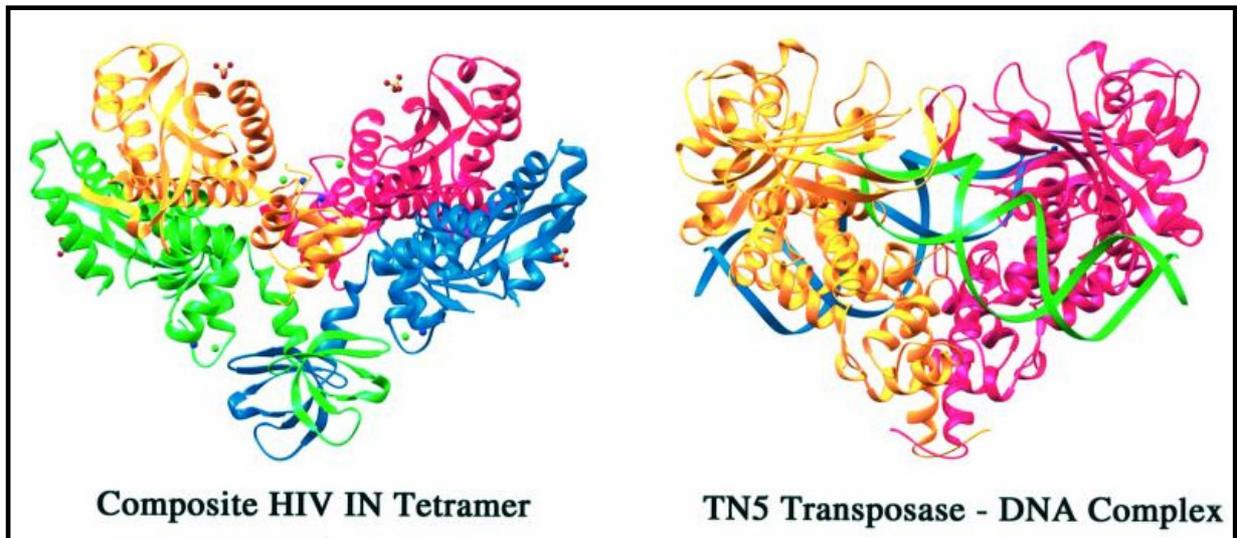
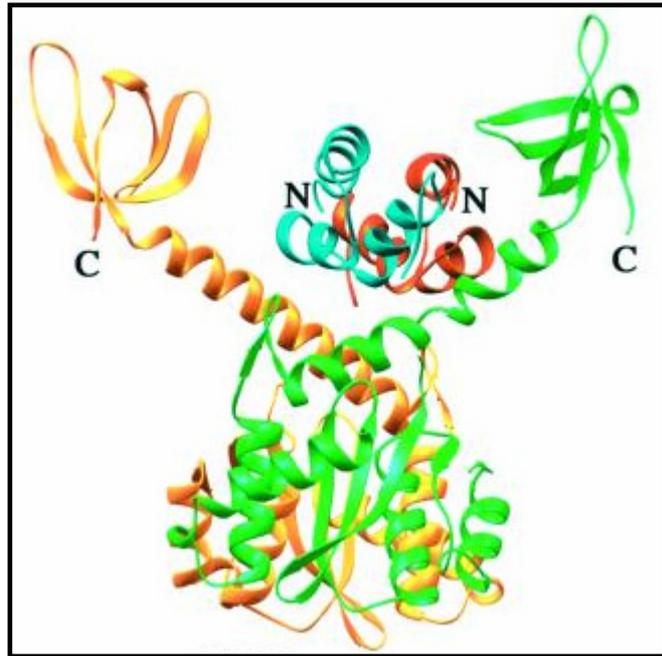
En l'absence de structure résolue à haute définition, seule une modélisation de la structure de l'intégrase entière et de ces formes multimériques peut être réalisée. Les modèles actuels de structure quaternaire sont réalisés à partir des structures cristallographiques des simples ou des doubles domaines de l'IN du VIH-1 et la majorité des études qui modélisent la structure du complexe IN / ADN sont réalisés sur le complexe IN / ADN viral / ADN cible qui serait le complexe responsable de l'intégration de l'ADN viral au sein du génome cellulaire. Les modèles sont également construits sur la base de données biochimiques d'interactions entre les différents domaines et avec les ADN viraux et cibles. Ces données ont été obtenues par pontage chimique, anisotropie de fluorescence résolue en temps, empreinte protéique, etc. En plus de ces données, les simulations s'appuient également sur l'homologie possible du complexe IN / ADN viral / ADN cible avec la structure cristallographique du complexe Tn5 Transposase / ADN transposon. Ces études ont pour but principal de déterminer une plateforme moléculaire pour réaliser des essais *in silico* de docking moléculaire entre l'IN et des petites molécules organiques en utilisant un modèle de cible plus significatif du point de vue fonctionnel.

La résolution de la structure des doubles domaines a permis d'émettre différentes hypothèses quand au réarrangement tridimensionnel des trois domaines entre eux. Après l'obtention de la structure du double domaine N-ter / CC, Wang et al. (Wang *et al.*, 2001) ont surimposé son CC (résidus 52-212) avec celui de la structure du double domaine CC / C-ter obtenue par Chen et al. (Chen *et al.*, 2000a). Les deux C-ter se situent à l'extrémité de l'hélice  $\alpha$  6 en n'ayant aucun contact entre eux ou avec les CC. Cette surimposition des CC montre un

**Figure 39 : Modélisation de la structure de l'IN à partir des doubles domaines.**

*D'après (Wang et al., 2001)*

Superposition des 2 structures par les CC (code PDB : 1EX4 et 1K6Y). Les domaines N-ter se positionnent sans encombrement stérique entre les deux domaines C-ter.



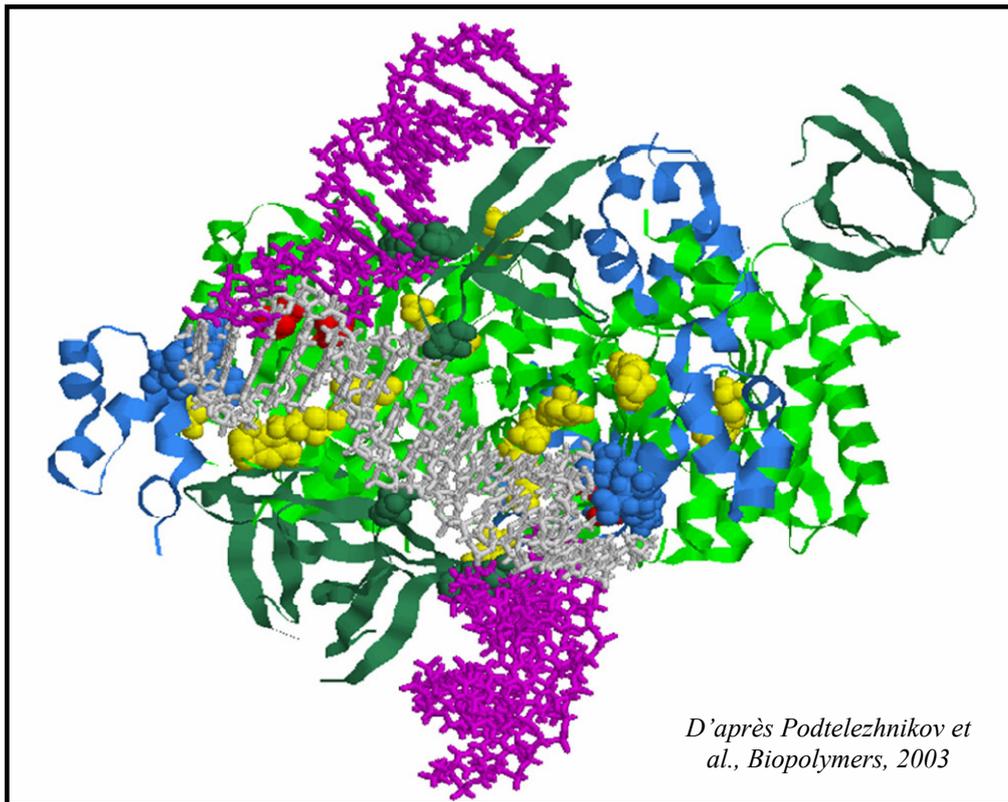
**Figure 40 : Modèle 1 du tétramère d'IN, comparaison avec le complexe Transposase Tn5/ADN**

Pour l'IN : les quatre sous-unités A, B,C et D sont respectivement en vert, jaune, bleu et magenta. Pour la transposase (code PDB : 1F3I) : les deux sous unités sont en jaune et magenta, les deux doubles brins d'ADN sont en bleu et vert. *D'après (Wang et al., 2001)*

positionnement du domaine N-ter exactement entre les deux hélices  $\alpha 6$  et ce sans aucune gêne stérique (cf. figure 39). Il y aurait même une possibilité de stabilisation du positionnement des deux hélices  $\alpha 6$  et donc des deux domaines C-Ter par des interactions avec les deux domaines N-Ter. La difficulté dans cette modélisation de l'arrangement spatial des trois domaines de l'IN du VIH-1 est importante suite à l'absence de certitudes quant au positionnement du domaine C-Ter par rapport au CC notamment en regard des différents arrangement structuraux des doubles domaines CC / C-ter des intégrases rétrovirales à ce jour cristallisées (IN de SIV<sup>50-293</sup> (Chen *et al.*, 2000b) et IN de RSV<sup>49-286</sup> (Yang *et al.*, 2000)). Ce qui ressort de ces comparaisons structurales c'est que la connexion entre les deux domaines reste extrêmement flexible et diffère en longueur, orientation et organisation au sein même de la famille des intégrases rétrovirales.

La majorité des modèles construits portent sur l'organisation tridimensionnelle du complexe responsable de l'intégration concertée. Comme nous l'avons vu précédemment, l'entité responsable de l'intégration concertée se compose au minimum d'un tétramère d'IN, des deux extrémités de l'ADN viral et de l'ADN cible. Les modèles développés mettent donc en présence un tétramère d'IN et deux ou trois ADN, deux de nature virale mimant chacun une extrémité de l'ADN viral et un mimant l'ADN cible. Plusieurs modèles sont disponibles. Seulement deux modèles seront présentés brièvement dans les paragraphes suivants.

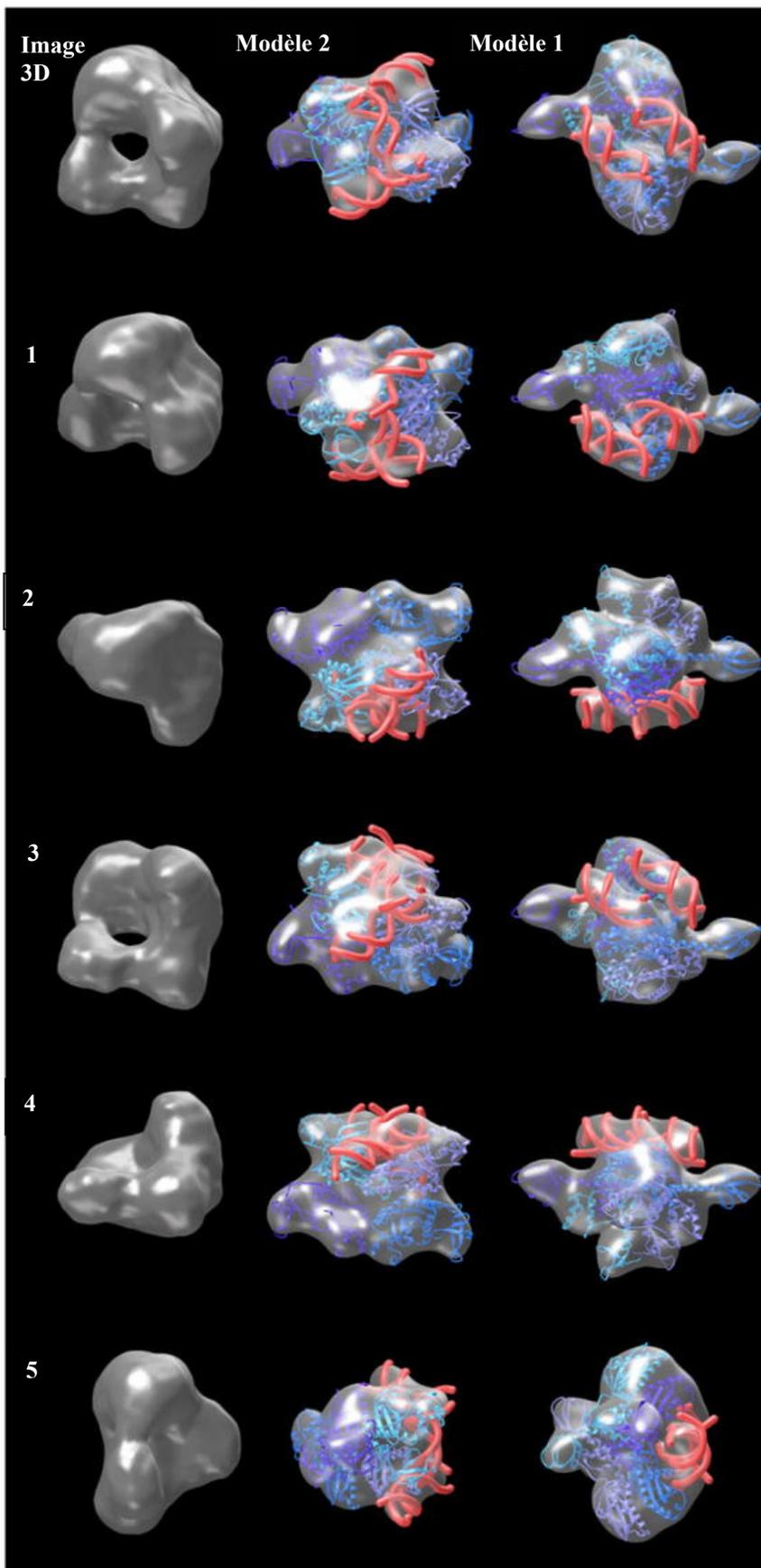
- Le premier modèle a été réalisé par Wang *et al.* (Wang *et al.*, 2001) à partir des deux structures des doubles domaines de l'IN du VIH-1 : IN<sup>1-212</sup> (code PDB : 1K6Y) et IN<sup>55-288</sup> (code PDB : 1EX4) qui sont superposés par leurs CC. L'arrangement du dimère de dimère a été construit sur le modèle de structure cristallographique obtenue pour la transposase Tn5 bactérienne (Davies *et al.*, 2000) (cf. figure 40). Cette corrélation entre les deux est possible puisque les deux protéines présentent une conservation de la structure du CC et un mécanisme de transfert de groupement phosphatidyl commun. Cette structure modélisée révèle une face concave pouvant présenter une surface potentielle d'interaction avec l'ADN cible. Le problème de ce modèle vient du fait que les deux sites actifs sont séparés d'environ 40 Å, cette distance est trop grande pour l'écart de 5 paires de bases observé lors de l'intégration concertée.
- Le deuxième modèle présenté a été construit par Podtelezhnikov *et al.* (Podtelezhnikov *et al.*, 2003). Il contient également un tétramère d'IN, les deux



**Figure 41 : Modèle 2 du tétramère d'IN, complexé avec les ADN viraux et cible**

En vert clair, les CC, en rouge et en modèle compact, les résidus D64 et D116, en vert, les domaines C-Ter. En modèle compact, les résidus S230 et R263 qui sont en contact direct avec l'ADN cible et l'ADN viral. En bleu, les domaines N-Ter. En modèle compact, les résidus F1, L2 et D3 qui sont en contact direct avec l'ADN cible. En gris et en violet, les fragments d'ADN cible et viraux. En modèle compact et en jaune, les résidus décrits comme étant en contact avec l'ADN viral d'après des études de pontage. *D'après (Podtelezchnikov et al., 2003).*

- extrémités de l'ADN viral modélisé sur 13 paires de bases tous deux déjà processé et d'un ADN cible de 17 paires de bases (cf. figure 41). Il prend en compte des données d'anisotropie de fluorescence résolue en temps (TFA) de l'IN et de ses oligomères ((Deprez *et al.*, 2000); (Leh *et al.*, 2000)). Au niveau structural, le modèle s'appuie sur les structures de domaines individuels (codes PDB pour le CC : 1BIS, le domaine C-ter : 1IHV et le domaine N-ter : 1WJA) et sur des données de pontage et d'empreinte protéique. L'arrangement des trois domaines entre eux est fait pour que la structure obtenue soit en accord avec les données de TFA, puis l'orientation des domaines est affinée en utilisant les données de pontage et d'empreinte protéique. Cette modélisation aboutit à un monomère plus compact que le précédent. Le dimère est construit en fonction des données de TFA avec une distance entre les deux sites catalytiques d'environ 45 Å. Enfin le tétramère est construit par arrangement de deux dimères avec plusieurs contraintes. Il faut que le transfert de brin soit réalisable avec un écart de 5 paires de bases sans trop perturber la conformation de l'ADN cible (distance entre les deux sites actifs d'environ 20 Å.). La symétrie de la réaction de transfert de brin doit être respectée. Pour cela, il faut que la structure présente un deuxième axe de symétrie dans l'organisation du tétramère centré sur les deux zones concernées par le transfert de brin. Enfin il faut que le modèle soit en accord avec les valeurs obtenues en TFA. Le modèle ainsi obtenu permet notamment d'expliquer les données expérimentales montrant que la mutation F185K bloque l'intégration correcte de l'ADN viral au sein de l'ADN cible *in vitro* et *in vivo* cette mutation stoppe la réplication virale. En effet, dans ce modèle, ce résidu se trouve à l'interface de tétramérisation. Il montre que l'introduction d'un résidu chargé rend impossible la formation du dimère de dimère. De plus, ce modèle explique, de part la position à l'interface des deux dimères qui comprend les deux domaines N-terminaux des monomères non vivo actifs, la tétramérisation de l'IN observée lorsque du  $Zn^{2+}$  se fixe au niveau du pseudo doigt de zinc (Zheng *et al.*, 1996). Tout en étant différent du premier modèle, ce modèle montre également une organisation structurale proche de celle observée pour la transposase Tn5.



**Figure 42 : Comparaison des surfaces du complexe 3D IN/ADN et des 2 modèles cités**

En rouge, l'ADN viral et cible, en bleus les monomères d'IN. A gauche : reconstitution 3D à 27 Å, A droite : modélisation de la surface avec une résolution de 27 Å pour les deux modèles. Les différentes vues sont réalisées en :

1 : rotation de 45° vers la gauche

2 : rotation de 90° vers la gauche

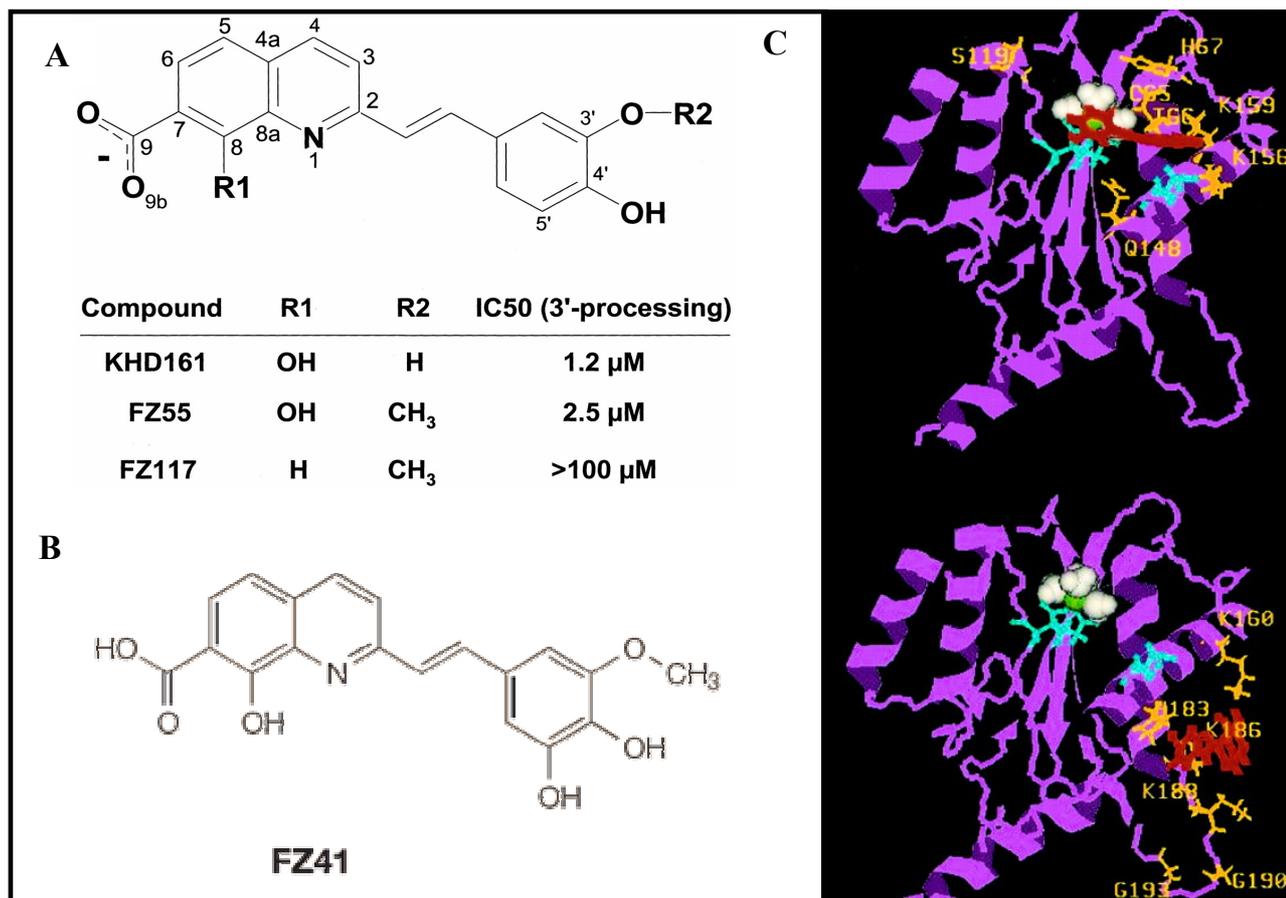
3 : rotation de 45° vers le haut

4 : rotation de 90° vers le haut

5 : rotation de 90° vers la droite

*D'après (Ren et al., 2007).*

En 2007, la structure de complexes IN / ADN structuré en Y (mimant l'intégration de l'ADN viral au niveau de l'ADN cible) fut obtenue avec une résolution de 27 Å (Ren *et al.*, 2007) par reconstruction tridimensionnelle du complexe IN / ADN à partir d'images obtenues par microscopie électronique (cf. figure 42). Cette résolution ne permet pas une précision dans le positionnement de l'ADN et des monomères d'IN pour déterminer la structure fine du complexe, mais elle permettrait d'orienter les études de modélisation futures du complexe IN / ADN compétent pour l'intégration. Cette image est composée d'un tétramère d'IN, déduction faite à partir du poids moléculaire du complexe de 128 kDa, qui présente quatre lobes. Chacun pourrait correspondre à un monomère d'IN. L'ensemble possède un diamètre total d'environ 50 Å. Trois des lobes forment une base triangulaire avec en son centre un canal d'environ 20 Å de diamètre. Le quatrième lobe se positionne entre deux des trois autres lobes et forme une extension d'environ 40 Å de haut. Les auteurs proposent que le canal central permettrait de loger l'ADN cible et que les deux extrémités de l'ADN viral soient liées par deux des quatre lobes. Ces données ont été comparées à différents modèles déjà proposés pour les complexes IN / ADN viral / ADN cible. Cette image du complexe met en évidence une asymétrie inattendue pour le complexe IN / ADN qui se retrouve dans beaucoup de complexes protéines / ADN dont ceux obtenus avec des topoisomérases et des polymérases. Aucun des modèles qui ont été réalisés jusqu'à présent ne se superpose à cette image (cf. figure 42). En effet, chacun présente une symétrie d'ordre deux que cette reconstruction d'image ne montre pas. Selon les auteurs, ceci pourrait être du à l'existence d'une flexibilité d'une certaine envergure au niveau du complexe IN / ADN.



**Figure 43 : Les styrylquinolines : inhibiteurs de type InBI**

**A** : Formule d'une styrylquinoline type avec sur le plan inférieur trois composés avec leur IC<sub>50</sub> sur l'activité de 3'processing sur l'IN du VIH-1

**B** : Formule semi-développée de FZ41

**C** : Docking sur le CC pour étudier les sites de fixation potentiels des styrylquinolines (ici le composé KHD161). 3 sites différents : le site préférentiel est positionné en partie dans le site actif en haut. Les deux autres pourraient être des sites de fixation de l'ADN cible.

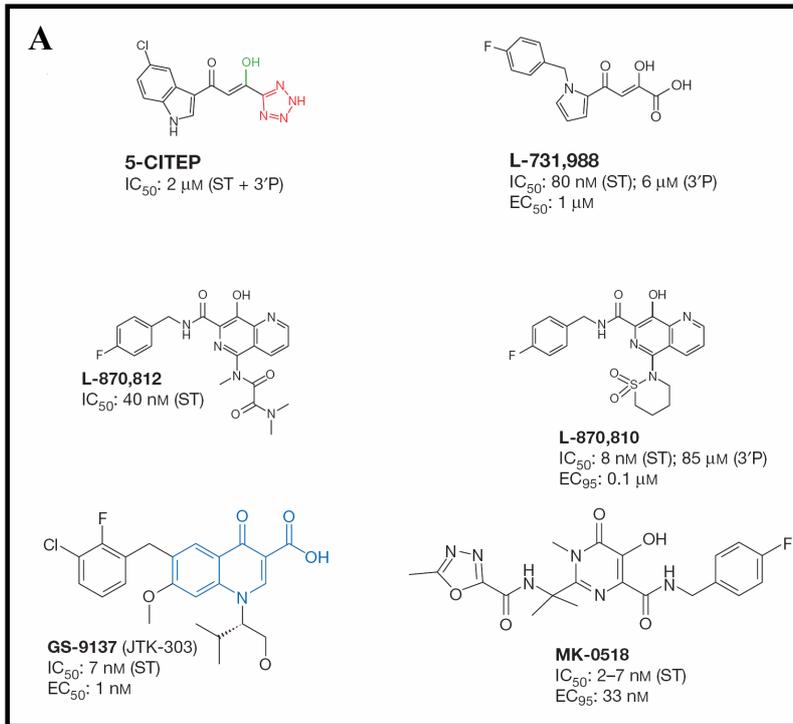
*D'après (Deprez et al., 2004)*

## **II.D. Les inhibiteurs de l'intégrase du VIH-1**

Plusieurs familles d'inhibiteurs de l'IN ont été identifiées. Leur classification repose principalement sur leur mode d'inhibition. Ils se divisent en trois classes : les inhibiteurs de la fixation de l'ADN cible (ou du transfert de brin), les inhibiteurs de la fixation de l'ADN viral (ou du 3'-processing) et les inhibiteurs modifiant l'état oligomérique de l'IN. La recherche d'inhibiteurs de l'IN du VIH-1 commence par un « screening » *in vitro* utilisé en routine. Les trois activités enzymatiques sont utilisées lors de ces tests. Pour étudier le transfert de brin indépendamment du 3'-processing, le substrat utilisé est de l'ADN pré-processé. La désintégration est notamment utilisée pour déterminer le site d'action de la drogue. Le CC pouvant catalyser cette réaction, il est alors possible de déterminer si le site catalytique est modifié par la molécule inhibitrice, ou si son action se situe à un autre niveau. Lorsqu'*in vitro* une molécule se révèle être active contre l'IN, son efficacité est testée sur des cultures cellulaires infectées par le VIH-1. Quatre critères sont importants pour montrer que l'IN est bien la cible de la drogue *ex vivo*. Dans un premier temps, des expériences à différents temps d'incorporation de la drogue sont réalisées et ils doivent montrer que sa fenêtre d'action est comprise entre la transcription inverse et la maturation (entre 4 et 16 heures après l'infection). Les cellules infectées traitées doivent également présenter une accumulation de cercles 2-LTR et une diminution du nombre de provirus intégrés dans le génome cellulaire. Les mutations de résistance développées par les virus doivent être situées au niveau de l'IN. Enfin, l'efficacité inhibitrice de la drogue doit être amoindrie lorsqu'elle est testée sur une IN qui porte les mutations mises en évidence précédemment.

### ***II.D.1. Les inhibiteurs de la fixation de l'ADN viral dit « InBI »***

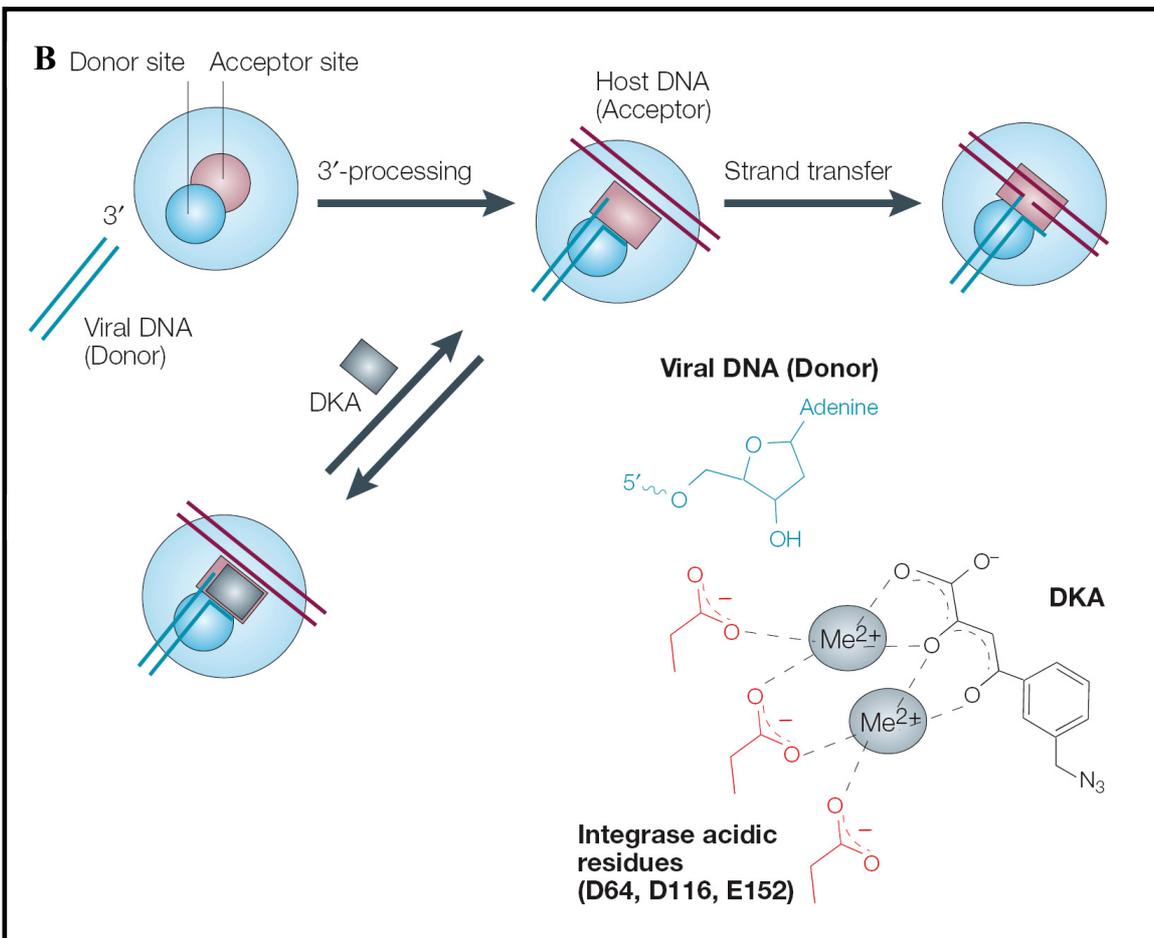
Ces inhibiteurs inhibent l'interaction de l'IN avec l'ADN viral. Même si leur mécanisme d'action *ex vivo* n'est pas encore très bien établi, ces molécules sont actives contre le 3'-processing qui est l'étape catalytique suivant la fixation de l'ADN viral (Bonnenfant *et al.*, 2004). Le développement de ce type d'inhibiteur est intéressant puisqu'il permet de bloquer la maturation de l'ADN viral qui est la première étape du processus d'intégration de l'ADN viral. La molécule type de ce type d'inhibition est FZ41, une molécule de la famille des styrylquinolines (cf. figure 43 A et B).



**Figure 44 : Les inhibiteurs du transfert de brin ou InSTI**

**A** : Formules de quelques InSTI caractéristiques : les diketo acides : 5-CITEP, L-731,988, L-870,812, L-870,810 ; les dérivés des quinoléines (groupement en bleu) : GS-9137 ou Elvitegravir et les dérivés des carboxamides : MK-0518 ou Raltegravir accepté par la FDA.

**B** : Mécanisme d'action des InSTI dérivés des diketo acides. Ils complexeraient le ou les cations divalents du site actif et bloqueraient ainsi toute activité de transfert de brin. *D'après (Pommier et al., 2005)*



*In vitro*, ces molécules agissent principalement en tant qu'inhibiteurs compétitifs de la fixation de l'ADN viral. Il faut noter que ces molécules ne possèdent pas la capacité de dissocier le complexe préformé IN / ADN viral. Des expériences de docking moléculaire (Deprez *et al.*, 2004) ont mis en évidence trois sites d'interaction forte des styrylquinolines avec l'IN. Deux de ces sites pourraient être responsables d'une action inhibitrice compétitive de la fixation de l'ADN viral et de l'ADN cible. D'ailleurs, outre l'étape de 3'-processing, ces composés peuvent également inhiber le transfert de brin mais avec une efficacité moindre que pour le 3'-processing (Deprez *et al.*, 2004) (cf. figure 43 C). Le troisième site ne contient pas de résidus impliqués dans la fixation de l'ADN mais serait davantage impliqué dans les interactions entre l'IN et d'autres partenaires protéiques. D'ailleurs, *ex vivo*, il a été montré que ces inhibiteurs pouvaient interférer avec (i) la fonction du complexe de préintégration au cours des étapes précoces du cycle viral et notamment lors de l'étape de transcription inverse (Bonnenfant *et al.*, 2004) et (ii) l'import nucléaire (Mousnier *et al.*, 2004). Il est donc possible que le troisième site joue un rôle dans le mécanisme d'action de ces drogues au niveau cellulaire.

Les mutations des virus résistants sélectionnés par FZ41 sont toutes portées par l'IN. Plusieurs mutations différentes sont introduites dans la séquence primaire de l'IN : la mutation C280Y et la double mutation V165I / V249I. Ceci conforte que l'IN est bien ciblée par les styrylquinolines *ex vivo* même si des processus, autres que l'intégration à proprement parler, peuvent être touchés. Il est intéressant de noter que le résidu V165 appartient au troisième site de fixation des styrylquinolines et que les virus résistants aux autres composés antirétroviraux (inhibiteurs de la RT) ou aux autres classes d'inhibiteurs de l'IN (les InSTI) sont sensibles à ces composés (Bonnenfant *et al.*, 2004). Ceci montre que les spectres des mutations de résistance générés par les autres composés antirétroviraux des phases précoces du cycle de réplication du VIH-1 ne possèdent aucun recouvrement avec celui généré par les styrylquinolines.

### ***II.D.2. Les inhibiteurs du transfert de brin dit « InSTI »***

Ces inhibiteurs seraient des compétiteurs de l'ADN cible. Ces inhibiteurs appartiennent à la famille des diketo acides ou à des dérivés plus ou moins éloignés comportant le motif structural  $\beta$ -hydroxyl-cétone (cf. figure 44 A). Ils agiraient en se positionnant à l'interface entre l'ADN cible et le site catalytique. Ces composés se fixent sur

le complexe ternaire IN / ADN viral en présence du cation divalent (Grobler *et al.*, 2002). Ils ne se fixent pas sur l'IN seule. Le développement d'inhibiteur s'est focalisé d'une part sur les interactions entre la drogue et le site actif notamment en chélatant le ou les cations divalents du site actif ( $Mg^{2+}$  ou  $Mn^{2+}$ ) par l'intermédiaire de son groupement  $\beta$ -hydroxyl-cétone (cf. figure 44 B) et d'autre part sur les interactions des groupements associés avec l'environnement du site actif et notamment une poche hydrophobe déterminée par la boucle flexible (Chen *et al.*, 2008).

Toute cette recherche a abouti sur différentes molécules actives et étudiées en test cliniques dérivées des diketo acides comportant des groupements carboxamide naphthyridine. Les mutations sélectionnées par les diketo acides sont situées dans le CC et dans la région de la triade catalytique. Elles ont les positions suivantes : N155S et la double mutation T66I / S153Y. Les molécules considérées comme de deuxième génération, c'est-à-dire les dérivés des diketo acides, vont induire des résistances croisées avec les diketo acides, mais également d'autres mutations de résistance. Parmi les molécules les plus étudiées, possédant le même mécanisme d'inhibition que les diketo acides, notons le composé L-870,810, un carboxamide naphthyridine, qui va induire deux types de mutations de résistance : la N155S, comme les diketo acides et une double mutation F121Y / T125K. Ces différences de mutations de résistances montrent un changement partiel de la zone d'interaction entre la drogue et l'IN (Hazuda *et al.*, 2004).

En plus des dérivés des diketo acides, d'autres type de molécules se sont révélées être inhibitrices du transfert de brin (cf. figure 44 A). Parmi ces composés, un dérivé de l'acide carboxylique dihydroquinoline, l'Elvitegravir, est en phase clinique I et II. Il possède le même type de mécanisme d'inhibition. Les mutations de résistance apparue en monothérapie sont toutes aussi situées dans le site actif, mais positionnés au niveau de résidus différents. Les mutations permettant une bonne résistance à l'Elvitegravir sont T66I, E92Q et dans une moindre mesure Q146P et S147G toutes deux situées dans la boucle flexible du CC (résidus 140 à 152) (Shimura *et al.*, 2008). Depuis novembre 2007, le Raltegravir, un dérivé des carboxamides, a été approuvé par la FDA (Food and Drug Administration). Les mutants de l'IN sélectionnés par cette dernière molécule sont N155H, E157Q, la double mutation G140S/Q148H, toutes situées dans la région de la boucle flexible et E92Q que l'on retrouve également pour l'Elvitegravir (Malet *et al.*, 2008).

### ***II.D.3. Les autres classes d'inhibiteurs de l'intégrase***

Parmi ces autres classes d'inhibiteurs, les « shiftides » occupent une place non négligeable. Cette classe d'inhibiteurs représente une approche différente dans la recherche d'inhibiteurs de l'IN. En effet, ces macromolécules polypeptidiques modulent l'état oligomérique de l'IN en forçant l'IN compétente pour le 3'-processing à passer sous une forme multimérique non compétente pour cette étape de maturation de l'ADN viral indispensable. C'est donc un mécanisme d'inhibition allostérique, qui présente l'avantage de n'avoir *a priori* aucune cross-résistance possible avec les InBI et les InSTI. Ces peptides peuvent être issus de l'IN elle-même comme INH5 qui correspond à son hélice  $\alpha$  5 (résidus 167 à 187). Ce peptide induirait, par une interaction avec l'hélice  $\alpha$  1, une dissociation des multimères en monomères (Maroun *et al.*, 2001). Les peptides peuvent également être issus d'interactants de l'IN comme LEDGF/p75. L'IN interagit avec ces peptides sous forme tétramérique, et devient incompétente pour la fixation à l'ADN. Il en résulte une inhibition de l'activité catalytique de l'IN *in vitro* et un blocage de la réplication virale en culture cellulaire (Hayouka *et al.*, 2007). La même équipe a montré également que des peptides issus de Rev peuvent également inhiber l'IN par le même mécanisme de transition état multimérique compétent à tétramère incompétent (Hayouka *et al.*, 2008).

Parmi ces autres classes d'anti-IN, des oligonucléotides ont également montré une activité inhibitrice *in vitro*. Une famille d'oligonucléotides formant des quartets de guanosines, dont AR177 (le Zintevir) (Mazumder *et al.*, 1996), comprend d'excellents inhibiteurs de l'IN. Cependant, l'étude des mutations virales sélectionnées par la molécule montre que son action antirétrovirale se situe *in vivo* au niveau de l'attachement viral. La molécule agirait sur SUgp120 (Cherepanov *et al.*, 1997). Cette action sur une autre étape du cycle réplcatif pourrait être symptomatique d'une absence de pénétration de la molécule dans les cellules.

Une deuxième famille d'oligonucléotides est actuellement étudiée. Ce sont des oligonucléotides composés de 11 bases modifiés au niveau de leur extrémité 3' par un groupement hydrophobe, qui peut être l'acridine ou l'éosine. Cette modification apporte à l'ADN un groupement hydrophobe permettant d'une part de faciliter le passage de la membrane plasmique lors des futurs tests *ex vivo* et d'autre part d'interagir avec la poche d'hydrophobicité également utilisée par les InBI et les InSTI. *In vitro*, ces oligonucléotides présentent un mécanisme d'inhibition compétitif du 3'-processing. Leur mécanisme est

quelque peu différent de celui employé par les styrylquinolines puisque ces oligonucléotides modifiés sont capables de dissocier les complexes IN / ADN viral préétablis (Pinskaya *et al.*, 2004). Des études sur le mécanisme d'action de ces oligonucléotides modifiés *ex vivo* sont actuellement en cours.

# **PRESENTATION DE LA** **THESE ET OBJECTIFS**

Nous avons pu voir tout au long de cette introduction que l'intégration de l'ADN du VIH-1 dans le génome des cellules hôtes est une étape obligatoire du cycle de réplication du virus. Le processus d'intégration est catalysé en deux temps par l'IN qui se fixe aux extrémités de l'ADN viral et l'intègre de manière covalente dans le génome cellulaire. Actuellement, parmi tous les composés antirétroviraux mis sur le marché, seul le Raltegravir, et ce seulement depuis décembre 2007, est dirigé contre cette étape. A ce jour, il existe des virus résistants à chacun des traitements proposés. Seule leur combinaison permet d'une part de retarder l'apparition des virus multirésistants et d'autre part de diminuer les doses administrées, grâce à des effets synergiques entre certains composés.

De ce fait, la recherche de nouveaux inhibiteurs et de nouvelles cibles d'action peut être vue comme une course contre le virus, ce dernier de part sa grande malléabilité génétique ayant un avantage non négligeable. Pour trouver de nouveaux composés actifs, la compréhension des étapes du cycle de multiplication potentiellement, inhibée par une thérapie, est essentielle. Il faut donc réaliser, en parallèle d'un criblage des molécules déjà connues, une étude approfondie des mécanismes mis en jeu au cours de l'étape de l'intégration du provirus du VIH-1. Leur compréhension permettra d'accélérer la recherche et le « design » de nouvelles molécules actives.

Ce travail s'est attaché à comprendre les paramètres qui régissent la cinétique de la première étape de l'intégration, le 3'-processing. D'après les données structurales, l'IN doit réaliser l'intégration concertée sous une forme multimérique minimale d'ordre quatre. L'état multimérique nécessaire pour la catalyse du 3'-processing, quant à lui, reste encore flou : est-ce que cette première étape nécessite le même état multimérique, ou un ordre inférieur est-il suffisant ? Afin de déterminer l'état multimérique de l'IN du VIH-1 responsable du 3'-processing, nous avons donc caractérisé les propriétés catalytiques et hydrodynamiques de l'IN du VIH-1 en présence d'oligonucléotides fluorescents mimant l'extrémité de l'ADN viral en utilisant l'anisotropie de fluorescence. Nous avons d'une part mis au point un test d'activité de 3'-processing de l'IN qui utilise l'anisotropie de fluorescence statique et d'autre part réalisé une caractérisation des complexes IN / ADN catalytiquement actifs par anisotropie de fluorescence résolue en temps (TFA). L'analyse du déclin multi-exponentiel de l'anisotropie de fluorescence par la méthode du maximum d'entropie permet de modéliser le déclin mesuré comme une somme d'exponentielles où chacune est associée à un temps de corrélation. Le temps de corrélation le plus long est relié au volume hydrodynamique de la

molécule ou du complexe observé. Sa détermination renseigne sur la taille de l'édifice moléculaire en question.

L'IN du VIH-1 possède un certain nombre de particularités comme une cinétique de catalyse lente et des conditions d'activité où la concentration d'enzyme doit être supérieure à celle de son substrat pour détecter une activité. La réaction de 3'-processing est spécifique de la séquence de l'ADN viral, notamment au niveau du CAGT terminal. Cependant, aucune interaction spécifique IN / ADN viral n'a été, à ce jour, mise en évidence. Il faut également noter que l'IN du VIH-1 est une protéine très peu soluble. De ce fait, souvent se pose la question de savoir si certaines caractéristiques de l'IN du VIH-1 sont des propriétés intrinsèques aux IN rétrovirales, ou bien sont seulement dues à la propension agrégative de l'enzyme. Afin de différencier les caractéristiques intrinsèques de l'IN de celles dues à son état agrégé, nous avons étudié l'IN de PFV-1 (Primate Foamy Virus de type 1, un spumavirus) qui présente un profil de solubilité plus clément. Nous avons alors pu réaliser un marquage enzymatique de l'IN de PFV-1 avec un fluorophore de petite taille (le TAMRA), qui nous a permis de réaliser des expériences de FCS et de FRET. La FCS est une technique complémentaire de la TFA. Elle donne des temps de diffusion translationnel de molécules fluorescentes, qui sont reliés au volume hydrodynamique de la molécule ou du complexe observé. Le FRET réalisé entre l'IN de PFV-1 et des oligonucléotides mimant l'extrémité des LTR viraux nous a permis de tester la spécificité de reconnaissance de l'ADN viral par l'IN de PFV-1.

La réaction de 3'-processing nécessite la présence des trois domaines structuraux de l'IN du VIH-1 alors que le CC est suffisant pour catalyser la coupure endonucléolytique. Les domaines N-ter et C-ter de l'IN jouent donc un rôle essentiel dans l'activité de 3'-processing. Différents résultats montrent que la structuration du domaine N-ter par la fixation d'un atome de zinc, au niveau d'un pseudo-doigt de zinc, est primordiale pour l'activité 3'-processing (Lee *et al.*, 1997). Les différentes activités catalysées par l'IN nécessitent la présence d'un cofacteur cationique ( $Mg^{2+}$  ou  $Mn^{2+}$ ) au niveau du site actif. Ces deux cations ne sont pas équivalents. Selon le cation utilisé *in vitro*, les activités de 3'-processing peuvent être différentes, par exemple lorsque le domaine N-ter est déstructuré. Nous avons utilisé le DIBA-1, qui est un éjecteur de zinc, afin de déterminer l'implication du domaine N-ter selon le contexte cationique dans les processus structuraux et catalytiques menant au 3'-processing.

# **MATERIEL ET** **METHODES**

# ***I. Protocole de purification de l'intégrase et tests d'activités in vitro***

## **I.A. Clonage et mutagenèse de l'intégrase**

L'IN du VIH-1 utilisée pendant mes travaux de thèse est une IN recombinante exprimée dans un système hétérologue bactérien. Toutes les constructions réalisées durant ce travail de thèse sont issues de la même construction initiale : viral de l'IN HBX2 du VIH-1 introduit dans le plasmide pET-15b (Leh *et al.*, 2000). Ce système permet de produire une IN contenant au niveau de l'extrémité N-Ter une étiquette 6-Histidine qui permet une purification en une seule étape décrite dans la partie suivante.

### ***I.A.1. Mutations ponctuelles sur l'intégrase du VIH-1***

Dans la troisième partie de cette thèse, différentes mutations ponctuelles ont été introduites dans la séquence de l'intégrase sauvage. Elles ont toutes eu pour but de muter un résidu cystéine en sérine. La première est la mutation C280S. Elle a été réalisée à l'aide du kit de biologie moléculaire QuikChange® II (Site directed Mutagenesis Kit de chez Stratagène). Ce système permet de réaliser des changements de résidus par une seule étape de PCR à l'aide d'oligonucléotides portant la mutation ponctuelle voulue.

Deux oligonucléotides complémentaires de séquence : 5'-CAG-ATG-GCA-GGT-GAT-GAT-TCA-GTG-GCA-AGT-AGA-CAG-GAT-3' ; ont été synthétisés (Eurogentec, Belgique). Le codon muté dont la séquence originelle était TGT (cystéine) est en gras dans la séquence. Nous avons suivi le protocole suivant pour chacune des mutations. Le mélange réactionnel (50 µL final) était composé de 5µL de tampon de réaction 10X, 20 ng de plasmide purifié à muter, de 125 ng final de chacun des deux oligonucléotide, 1 µL du mélange de dNTP, 2,5 unités de Pfu ultra (polymérase ADN-ADN dépendante possédant un taux d'erreur très faible) et pour finir le volume réactionnel est complété par de l'eau MilliQ stérile.

Les cycles de PCR ont été réalisés selon le schéma suivant :

30 sec à 95°C	}	16 cycles
30 sec à 95°C		
1 min à 55°C		
7 min à 68°C (longueur plasmide ≈ 7 kb)		

7 min à 68°C, puis à 4°C pendant 1h

10 unités de Dpn I sont ensuite ajoutées afin de digérer le plasmide parental ne portant pas la mutation. La réaction se passe à 37°C pendant 1 heure.

Des bactéries compétentes XL1 blue ou XL10 gold sont transformées avec les volumes suivants : 1 µL de mélange de PCR (pas plus car la transformation est inhibée par la quantité d'oligonucléotides et les fragments d'ADN plasmidique parental) pour 50 µL de bactéries compétentes. La transformation est réalisée selon le protocole concernant chaque souche utilisée avec un choc à 42°C de 45 secondes pour les XL1 Blue et de 30 secondes pour les XL10 Gold. Les bactéries transformées sont étalées sur boîtes contenant de l'antibiotique et mises sur la nuit à 37°C ou sur 2 jours à température ambiante. Les clones obtenus sont, après préparation de miniprep de plasmides, séquencés afin de vérifier la présence de la mutation voulue et l'absence de toute autre mutation ponctuelle dues à une erreur de la Pfu ultra.

Deux autres cystéines ont été également mutées :

C56S avec l'oligonucléotide suivant : 5'-CAT-GGA-CAA-GTA-GAC-TCT-AGC-CCA-GGA-ATA-TGG-CAG-CTA-3', où le codon en gras remplace le codon TGT codant pour la cystéine initiale.

C65S avec l'oligonucléotide : 5'-GA-ATA-TGG-CAG-CTA-GAT-TCT-ACA-CAT-TTA-GAA-GGA-AAA-G-3', où le codon en gras remplace également le codon TGT qui codait pour la cystéine initiale.

Ces mutations ont été introduites dans un plasmide codant pour l'IN du VIH-1 portant déjà la mutation C280S. Le triple mutant C56, 65, 280S a été obtenu à partir du double mutant C65, 280S.

### ***1.A.2. Délétions sur l'IN du VIH-1***

Pour réaliser les délétions d'un domaine entier de l'IN du VIH-1, nous avons amplifié la séquence de l'IN C280S, obtenue par mutagenèse dirigée, délétée du domaine non désiré dans la construction finale par PCR. Ce fragment de PCR a été ensuite introduit par ligation dans un système d'amplification (p-GEMT easy vector). Ce système permet de séquencer le fragment de PCR afin de vérifier l'absence de mutations ponctuelles introduites lors de l'étape de PCR. Le plasmide, ainsi vérifié, est ensuite digéré par deux enzymes de restrictions : NdeI et BamHI. NdeI clive au niveau d'un site de restriction (CAT-ATG) qui est positionné au

niveau de l'ATG initiateur. Ce site permet une insertion du fragment de l'IN dans pET-15b au niveau du bon cadre de lecture. De plus, dans le système pET-15b que nous avons utilisé, ce site de restriction est situé en aval du tag 6-Histidine, nécessaire pour notre protocole de purification. Le site de restriction de BamHI est situé en aval du codon stop de la séquence codant pour l'IN.

Pour réaliser les mutants délétés du domaine N-Ter, un oligonucléotide de séquence : 5'-**CAT-ATG-CAT-GGA-CAA-GTA-GAC-TG**-3' a été synthétisé. Le site de restriction de NdeI est en gras et la séquence complémentaire de viral codant pour l'IN à partir du résidu Met 50 est soulignée. La deuxième amorce utilisée pour la PCR est soit T7 ter de séquence : 5'-TAT-GCT-AGT-TAT-TGC-TCA-G-3', situé en aval du codon stop de l'intégrase entière, cette PCR aura pour construction finale IN<sup>50-288</sup>, soit un oligonucléotide insérant un codon stop après le résidu Leu 213. La séquence de cet oligonucléotide est 5'-**CTA-TAA-TTC-TTT-TAG-TTT-GTA**-3' (le codon stop introduit est en gras et la séquence complémentaire de viral de l'IN du VIH-1 des résidus 213 à 207 pour la construction IN<sup>50-213</sup> est soulignée). Pour le mutants délétés uniquement du C-ter, IN<sup>1-213</sup>, nous avons fait une PCR entre l'oligonucléotide précédant et l'oligonucléotide T7 prom de séquence 5'-TAA-TAC-GAC-TCA-CTA-TAG-G-3' situé en amont du site de restriction de NdeI.

Ces PCR permettent donc d'obtenir des inserts contenant la séquence codant de l'IN du VIH-1 tronquée. Après vérification de la séquence, le plasmide pGEMT contenant l'insert est digéré par NdeI et BamHI. Ce fragment est purifié sur gel d'agarose 1%, puis inséré par ligation dans un plasmide pET-15b lui-même prédigéré par NdeI et BamHI. Pour la construction IN<sup>50-288</sup>, il n'était pas possible de séparer le plasmide superenroulé pGEMT contenant l'insert de l'insert après digestion par NdeI et BamHI, pour s'affranchir de cette contamination, nous avons utilisé un plasmide pET-15b dont le gène de résistance à l'ampicilline a été remplacé par le gène de résistance à la kanamycine. Ce changement d'antibiotique nous a permis d'obtenir cette construction.

## I.B. Expression

Le protocole utilisé pour réaliser l'expression hétérologue de l'IN du VIH-1 a subi quelques modifications par rapport à celui initialement publié par Leh et al. (2000) (Leh *et al.*, 2000).

Une pré-culture des bactéries contenant le plasmide d'expression voulu est réalisée sur la nuit à 37°C. Le milieu de culture est composé de LB (10g de NaCl, 10g de Bactotryptone et 5g d'extrait de levure dans 1 L d'eau) additionné de l'antibiotique de sélection (Ampicilline 100 µg/mL ou Kanamycine 30µg/mL). Lorsque le plasmide est introduit dans la souche Rosetta ou BL21 DE3 Codon plus, il faut ajouter du Chloramphénicol (30 µg/mL).

10 à 20 mL de la pré-culture sont ajoutés à 500 mL de LB contenant l'antibiotique de sélection (Ampicilline ou Kanamycine). Le Chloramphénicol n'est plus nécessaire. La culture est ensuite maintenue à 37°C, sous une agitation de 180 rpm pour une bonne oxygénation des bactéries. Lorsque la  $DO^{600nm}$  est comprise entre 0,6 et 0,8, l'IPTG est ajouté à une concentration finale de 0.25mM pour induire la production. L'expression peut alors être réalisée à deux températures différentes selon le profil de solubilité de l'IN recombinante : 37°C pendant 3 à 4 heures si la protéine reste soluble dans les bactéries ou 25°C pendant 6 à 7 heures si la protéine lorsqu'elle est produite est dirigée vers les corps d'inclusions. La culture bactérienne est ensuite centrifugée à 3500 rpm pendant 15 à 20 minutes, le surnageant est ôté et les culots sont congelés à -80°C.

## I.C. Protocole de purification

La composition des différents tampons est explicitée en annexe 1.

Les culots bactériens sont resuspendus dans le tampon de resuspension à hauteur de 25 mL pour 1 L de culture bactérienne environ. Les bactéries sont ensuite lysées avec la presse de French (5 à 6 passages) ou le sonicateur (fréquence de 60Hz, pulses de 9 secondes, pendant trois minutes deux fois avec un refroidissement dans la glace entre les deux cycles de sonication). Le lysat est ensuite centrifugé pendant 30 min à 10000 rpm à 4°C. Le surnageant est récupéré puis filtré avec des filtres de 0.2 / 0.8 µm.

Le filtrat est ensuite mis en contact des billes NiNTA préalablement rincées avec du Tampon A, puis avec du Tampon A additionné de 40mM d'imidazole. Il faut compter entre 2.5 et 5 mL de billes pour 1 L de culture bactérienne selon la quantité d'IN produite. Le mélange billes NiNTA / filtrat est maintenu sous une agitation douce à 4°C pendant 2h au minimum.

Le mélange est ensuite centrifugé à 3000 rpm pendant 10 min pour sédimenter les billes, le surnageant est prélevé puis conservé à 4°C (au cas où l'IN ne se soit pas fixée sur les billes). Les billes sont lavées quatre fois avec 40 mL de Tampon A additionné de 40 mM d'imidazole avec des centrifugations de 10 min à 3000 rpm pour faire sédimenter les billes entre chaque lavage. Lorsque le surnageant ne contient plus de protéines (absence de coloration bleu lors d'un test visuel en plaque au Bradford (190 µL Bradford 1X et 10 µL surnageant ou billes), il faut vérifier que la fraction des billes contient des protéines. Les billes sont ensuite lavées quatre fois avec 40 mL de Tampon A additionné de 80 mM d'imidazole puis si nécessaire deux fois avec 40 mL de Tampon A additionné de 100 mM d'imidazole.

Lors du dernier lavage, les billes sont transférées dans des tubes Eppendorf de 2 mL et la première élution est réalisée avec un volume restreint de tampon d'élution (pour 1 mL de billes sèches, 0.5 mL de Tampon d'élution, le mélange est maintenu sous agitation douce pendant 20 minutes pour chacune des éluions. Les deuxième, troisième et quatrième éluions sont réalisées dans un volume plus important (entre 0.75 et 1 mL). La concentration protéique des éluats est vérifiée par la technique du Bradford et les fractions contenant les mêmes concentrations sont rassemblées.

Pendant les éluions successives, le boudin de dialyse est coupé à la longueur nécessaire estimée puis rincé trois fois 10 min avec de l'eau MilliQ, puis dans du Tampon A filtré. Le tampon de dialyse est mis à 4°C de manière à ce qu'il soit homogène en température avec les boudins de dialyse contenant l'éluat. Les boudins sont fermés à une extrémité après avoir rincé l'intérieur avec du Tampon A filtré puis remplis avec l'éluât à l'aide d'une pipette. Le boudin est ensuite fermé en laissant 1 cm d'air dans sa partie supérieure puis plongé dans le tampon de dialyse. La dialyse se fait sur la nuit à 4°C sous faible agitation. Une deuxième dialyse peut être réalisée le lendemain sur 4 h au minimum.

L'IN est récupérée puis dosée au Bradford. Elle est ensuite aliquotée en fraction de 20 à 100µL selon les expériences réalisées par la suite. L'IN est congelée rapidement dans de l'azote liquide ou dans de la carboglace additionnée d'éthanol. Les aliquots sont conservés à -80°C.

Dans le cadre de certaines expériences nécessitant des mélanges réactionnels particuliers, nous avons modifié ce protocole à différents niveaux. Pour les expériences réalisées en présence de DIBA-1, l'IN ne doit pas contenir d'agent réducteur. Le β-mercaptoéthanol n'est pas ajouté dans le Tampon A utilisé pour les lavages à 80 mM et 100 mM et les éluions. Le Tampon de dialyse n'en contient pas non plus. Les premiers tampons de lavage à 40 mM et dans le tampon de resuspension doivent en contenir car il prévient la formation des ponts disulfures intermoléculaires entre l'IN et des protéines bactériennes contaminantes. Trois dialyses successives sont réalisées pour éliminer tous les résidus de β-mercaptoéthanol. De plus, pour les expériences de dosage de l'éjection du zinc de l'IN par le DIBA-1, les tampons de dialyse ne contiennent pas de zinc. Trois dialyses sont également réalisées afin de diminuer au maximum les contaminations dues aux tampons. Dans les cas où l'IN précipite pendant la ou les dialyses. Sa solubilité peut être améliorée en abaissant le pH du tampon de dialyse à 6,5 et / ou en réalisant des paliers de concentration d'imidazole : 300mM, 150mM, puis 0mM. Enfin la dernière solution est d'estimer la concentration de l'IN avant dialyse et de la diluée de manière à ce qu'elle n'agrège pas. La dialyse peut également être remplacée par un échange de tampon progressif réalisé avec des amicons de cut-off de 10 kDa et avec un tampon de composition identique à celui de la dialyse.

## **I.D. Marquage enzymatique de l'IN avec un petit fluorophore**

Le marquage est réalisé sur l'IN (du VIH-1 ou de PFV-1) au niveau de son extrémité C-terminale sur une étiquette de séquence PKPQQFM spécifique pour la transglutaminase de foie de cochon d'inde.

Le marquage est réalisé lorsque l'IN est fixée sur les billes NiNTA par le tag Histidine de son extrémité N-Terminale, après les différents lavages en présence d'imidazole (l'imidazole doit être au préalable éliminé). Le marquage se fait par incubation à 4°C sur la nuit de l'IN en présence de 0.5 mM de TAMRA couplé à la cadavérine (Molecular Probes) avec une unité de transglutaminase (Sigma) dans un tampon de marquage composé de 20 mM Hepes, pH 7.5 et de 10 mM de Ca<sup>2+</sup>. Le TAMRA non transféré sur l'IN est ensuite éliminé lors de lavages. L'éluion est réalisée comme pour le protocole ci-dessus. L'IN subit alors une étape de purification supplémentaire afin d'éliminer le NaCl et l'imidazole. L'éluât est passé sur une colonne échangeuse de cation MonoS (Biorad), équilibrée avec un tampon contenant 20mM Hepes pH 7.5, 0.1 M NaCl. Avant de charger l'IN sur la colonne (débit : 1 mL/min), elle est diluée 10 fois (diminution de la force ionique). Un gradient de NaCl est appliqué (de 0.1 à 1 M). L'IN est éluée entre 450 et 550 mM de NaCl. Cette étape n'est possible que pour l'IN de PFV-1, car celle du VIH-1 de par sa faible solubilité agrège dans la colonne. Il est ensuite possible de concentrer l'IN de PFV-1 avec un système de filtres par centrifugation Amicon Ultra (Millipore). La dialyse n'est pas employée pour les concentrations d'IN de PFV-1 supérieures à 12 mg / mL, car à ces concentrations d'IN, elle précipite durant les étapes de dialyse lorsque la concentration de l'imidazole est inférieure à 250 mM.

La spécificité du marquage a été vérifiée par spectrométrie de masse et le taux de marquage a été déterminé par spectrophotométrie et par FCS. Il a été évalué à 50%. De plus, l'IN marquée ne présente pas de profil d'activité différent de celui observé pour l'IN non marquée (pour plus de détail voir la partie résultat de l'article 2 « Etude du mécanisme de reconnaissance intégrase / ADN : un mode d'interaction à l'ADN spécifique mis en évidence par une intégrase marquée avec une technique enzymatique »).

## **I.E. Tests d'activités**

Les tests d'activité de l'IN ont été réalisés avec deux méthodes différentes : un test utilisant l'anisotropie de fluorescence et un test utilisant des oligonucléotides marqués avec du  $^{32}\text{P}$  à leur extrémité 5' du brin d'ADN clivé. Les séquences des oligonucléotides sont toutes explicitées dans le matériel et méthodes de chacun des manuscrits des articles. Il en est de même pour les conditions expérimentales.

### ***I.E.1. Test utilisant l'anisotropie de fluorescence statique***

#### **I.E.1.a. Présentation de l'appareillage : le Beacon et l'Eclipse**

Les deux appareils suivant nous ont permis de mesurer l'anisotropie des molécules fluorescentes en solution :

Beacon ®2000 (PanVera, Madison, USA) : c'est un instrument de mesure de la polarisation ou de l'anisotropie de fluorescence statique d'échantillons liquides dont le volume peut varier de 100  $\mu\text{L}$  à 1 mL (cuves en verre non polarisant). La chambre de mesure est thermostatée (6°C à 65°C). Cet appareil programmable est piloté par ordinateur. Il est muni de filtres d'excitation et d'émission qui permettent de travailler avec des concentrations de fluorophores de l'ordre du nanomolaire. Les filtres utilisés permettent de travailler avec la fluorescéine qui est excitée à 495 nm et qui émet vers 515 nm.

Cary Eclipse Fluorescence Spectrophotometer (Varian) : c'est un spectrofluorimètre classique qui permet de mesurer l'intensité de fluorescence des échantillons dont les volumes sont compris entre 50  $\mu\text{L}$  et 1 mL. La chambre de mesure est thermostatée à l'aide d'une instrumentation externe (CC240, Hüber). L'Eclipse possède deux monochromateurs un pour l'excitation et un pour l'émission, qui permettent de travailler sur une gamme spectrale comprise entre 190 et 900 nm. Des polariseurs peuvent être ajoutés respectivement à l'excitation et à l'émission afin de mesurer les composantes parallèles et perpendiculaires de l'intensité de fluorescence, ce qui permet de calculer l'anisotropie de fluorescence statique des échantillons.

### I.E.1.b. Test d'activité

Pour les différentes expériences réalisées dans les cuves en verres ou en quartz, les contenants sont dans un premier temps traités avec une solution de BSA (Sigma A0281) à 20 mg / mL durant 20 minutes. La solution est ensuite ôtée. Ce prétraitement permet de limiter l'adsorption de l'IN sur les parois des tubes.

Les conditions d'activités seront décrites dans chacun des articles au niveau des légendes des figures, cependant même si les concentrations peuvent varier, la composition du mélange réactionnel comprend : le tampon Hepes 20 mM, pH 7, additionné de l'ADN processable, d'IN, de 10 mM de cation divalent ( $Mg^{2+}$  ou  $Mn^{2+}$ ), d'1 mM de DTT et de 50 mM de NaCl. Pour les tests utilisant le DIBA-1, toutes les expériences ont été menées avec une concentration finale de DMSO de 15 % en l'absence de DTT.

Un mélange équimolaire des deux brins complémentaires d'ADN est hybridé dans du tampon HEPES 20 mM, pH 7, avec 100 mM final de NaCl pour former l'ADN double brin que va cliver l'IN. La concentration finale en ADN est de 1  $\mu$ M. Le mélange est porté à une température de 85°C pendant 5 min. Les deux brins s'apparient spécifiquement par refroidissement lent du mélange à température ambiante pendant 30 minutes.

Les oligonucléotides sont purifiés lors de leur réception sur gel acrylamide de séquence de 12 à 15% acrylamide (selon le nombre de bases des oligonucléotides), puis passés sur colonnes NAP<sup>TM</sup>5 (Amersham) pour éliminer les sels. L'éluion est réalisée avec 1 mL d'eau MilliQ. Les oligonucléotides, après avoir été quantifiés au spectrophotomètre à 260 nm, sont ensuite séchés au Speed Vac Plus (SC110A, Savant) et repris dans tampon Tris HCl 20mM, pH 7,0 à une concentration comprise entre 50 et 100  $\mu$ M. Leur concentration est ensuite mesurée par absorbance à 260 nm. Ils sont conservés à -20°C.

Ce test s'appuie sur les propriétés additives de l'anisotropie de fluorescence. L'ADN, marqué avec la fluorescéine au niveau du T de son extrémité CAGT (pour l'IN du VIH-1) ou CAAT (pour l'IN de PFV-1), possède une anisotropie statique relative à sa taille et à sa séquence. Lorsque l'IN se fixe sur cet ADN, le complexe fluorescent observé possède alors un volume hydrodynamique plus élevé, et donc une anisotropie plus importante. Durant la fixation de l'IN sur l'ADN on observe donc une augmentation de l'anisotropie statique du mélange réactionnel. Lorsque la température passe à 37°C, l'IN catalyse la coupure endonucléolytique libérant ainsi dans le milieu réactionnel le dinucléotide GT (pour l'IN du

VIH-1) ou AT (pour l'IN de PFV-1) fluorescent. Ce dinucléotide, de part sa taille, possède une anisotropie statique très inférieure à celle du complexe IN / ADN et à celle de l'ADN seul. De ce fait l'anisotropie statique du mélange réactionnel chute proportionnellement à sa libération dans le milieu réactionnel. Le principe du test est décrit plus précisément dans le matériel et méthodes et les résultats de l'article 1 : « Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity ».

### ***1.E.2. Test d'activité utilisant les oligonucléotides radiomarqués***

#### Marquage des oligonucléotides en 5' par la T4-polynucléotidylkinase.

Le mélange réactionnel est constitué d'AT $\gamma$ P<sup>32</sup> (2.5  $\mu$ curie /  $\mu$ L final), de 2  $\mu$ L de tampon de réaction 10X, 10 unités de Kinase, de l'oligonucléotide processable (500nM final). Le volume réactionnel est complété à 20  $\mu$ L par de l'H<sub>2</sub>O MilliQ stérile.

La réaction est réalisée à 37°C pendant 1h / 1h30. Elle est stoppée par ajout d'1  $\mu$ L d'EDTA à 0.5 M, puis la kinase est inactivée par chauffage du mélange réactionnel à 65°C pendant 10 minutes. L'ADN marqué est ensuite hybridé avec 1  $\mu$ L d'oligonucléotide complémentaire à 10  $\mu$ M (concentration finale de 250 nM) en présence de 100 mM de NaCl. Le volume réactionnel de 40  $\mu$ L est complété avec de l'H<sub>2</sub>O MilliQ stérile.

Le mélange est chauffé à 85°C pendant 5 minutes puis l'hybridation est réalisée à température ambiante pendant 30 min. L'ADN est ensuite purifié (élimination de l'AT $\gamma$ P<sup>32</sup> et de l'EDTA) sur une colonne G25 (Amersham). A la fin de la purification, on vérifie la présence de radioactivité dans la phase recueillie.

#### La réaction de 3'processing.

Le mélange réactionnel de 20  $\mu$ L est composé de Tampon Hepes 20mM pH 6,8 final, de DTT à 1mM final (quand cela est possible), de DMSO ou de DIBA-1 dans du DMSO avec 15 % de DMSO final, 10 mM final de cation divalent (Mg<sup>2+</sup> ou Mn<sup>2+</sup>), X nM final d'IN et X nM final d'ADN db radiomarqué (les concentrations sont explicités pour chacune des figures). Enfin, l'H<sub>2</sub>O MilliQ stérile est ajoutée pour compléter le volume à 20  $\mu$ L.

Le 3'-processing est réalisé à 37°C. Tout changement dans ce protocole est indiqué dans chacune des légendes des figures concernées. La réaction est ensuite arrêtée avec 80  $\mu$ L d'un mélange contenant (volumes donnés pour 1 mL de cette solution stop) de l'EDTA à 62.5

mM final (12.5  $\mu$ L d'une solution à 0.5 M), de l'AcNa à 0.3 M final (100  $\mu$ L d'une solution à 3 M) et du glycogène à 0.1 mg / mL final (10  $\mu$ L d'une solution à 10 mg / mL).

Une extraction au phénol chloroforme permet d'éliminer l'IN. La phase organique située au fond du tube est prélevée puis jetée. La phase aqueuse surnageante contenant l'ADN est conservée. Les oligonucléotides sont précipités à l'éthanol (2,5 vol d'éthanol absolu / 1 vol d'extrait, les tubes sont placés une nuit à - 20°C, puis centrifugés à 13400 rpm pendant 30 min). Après l'élimination du surnageant, les oligonucléotides sont séchés puis repris dans du formamide / EDTA (20 mM final) de manière à avoir 10000 cpm /  $\mu$ L (le volume minimum de resuspension est de 5  $\mu$ L). Les oligonucléotides sont ensuite déposés sur gel acrylamide 12 % (40000 cpm au maximum par puits). La migration dure environ 1h15 pour une puissance de 80 W.

Le gel est ensuite séché. La révélation se fait avec une plaque phosphorescente qui est excitée par le rayonnement  $\beta$  du  $^{32}\text{P}$ . La lecture se fait avec le Storm 840 qui scanne la plaque et mesure l'intensité de la phosphorescence émise. La quantification se fait à l'aide du logiciel Image Quant.

## ***II. Présentation du principe de la fluorescence et des techniques utilisées***

Nous allons présenter rapidement le phénomène de fluorescence, l'anisotropie de fluorescence qui se décline en deux types de mesures selon la méthode d'acquisition : l'anisotropie résolue en temps et l'anisotropie statique. Ensuite nous présenterons succinctement les techniques de FRET et de FCS également utilisées (Bibliographie : Lakowicz, Principles of fluorescence spectroscopy, chapitre 1, 10, 11, 13 et 24, p1 à 24, 353 à 380, 383 à 408, 443 à 471 et 798 à 837).

### **II.A. Le phénomène de fluorescence**

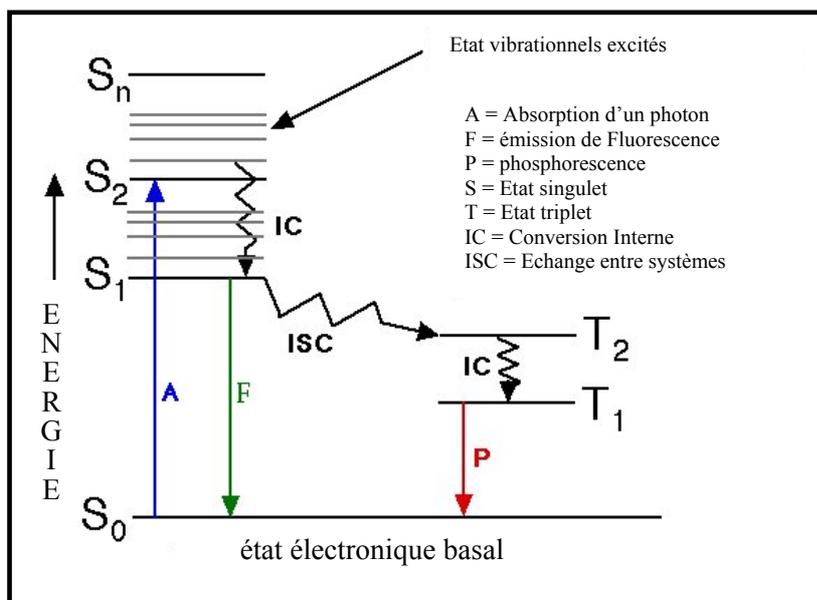
Lorsqu'une molécule est excitée, l'un de ses électrons passe du niveau d'énergie fondamental à un niveau d'excitation plus élevé. Ce changement d'orbitale est instable et l'électron retourne rapidement à son niveau d'énergie fondamental. Lorsqu'il y a fluorescence de la molécule, ce changement d'orbitale vers un niveau plus stable s'accompagne de l'émission d'un photon. L'excitation et l'émission sont usuellement illustrées par le diagramme de Jablonski (cf. figure 45). Lors du phénomène de fluorescence, le taux d'émission de la fluorescence est d'environ  $10^8 \text{ s}^{-1}$ , ce qui correspond à un temps de vie de fluorescence typique d'environ 10ns. Le temps de vie ( $\tau$ ) d'un fluorophore correspond à la moyenne du temps mis pour diminuer l'intensité initiale (à  $t = 0$ ) d'un facteur  $e$ .

Si l'on observe le diagramme de Jablonski (cf. figure 45), on peut remarquer que l'énergie de l'émission est plus basse que l'énergie d'excitation, ce qui se visualise par une émission à une longueur d'onde plus élevée que celle de l'excitation. C'est le décalage de Stokes (cf. figure 46). En fluorescence, deux données sont primordiales lorsque l'on étudie un fluorophore : le rendement quantique ( $Q$ ) et le temps de vie ( $\tau$ ). Ces deux valeurs sont calculées avec les formules suivantes :

$$Q = \Gamma / (\Gamma + k_{nr}) \quad (1)$$

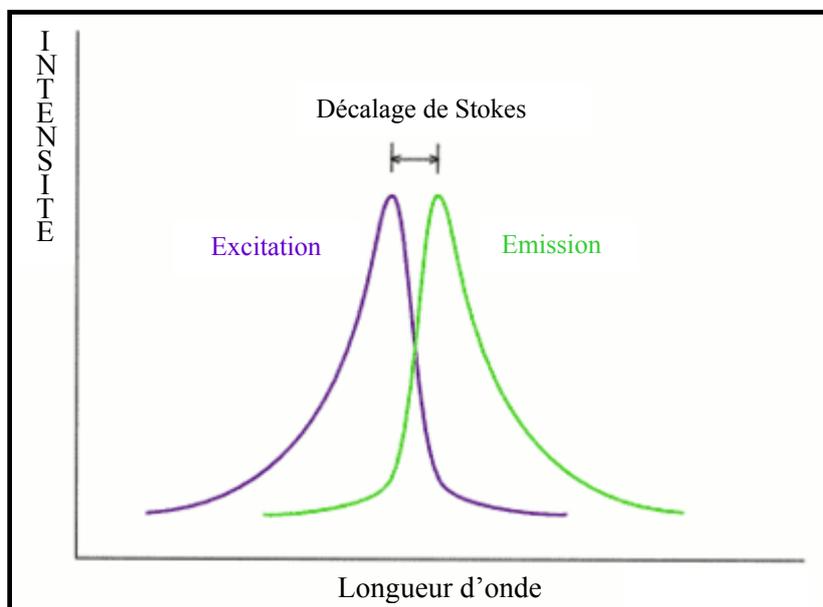
$$\tau = 1 / (\Gamma + k_{nr}) \quad (2)$$

Où  $\Gamma$  est le taux d'émission et  $k_{nr}$  la constante de temps non radiative du fluorophore.



**Figure 45 : Diagramme de Jablonski**

Description du phénomène de fluorescence par le diagramme de Jablonski



**Figure 46 : Superposition du spectre d'excitation et d'émission d'un fluorophore.**

Sur ce graphe, on peut voir le décalage de Stokes dû à la diminution de l'énergie du photon entre son adsorption et son émission.

## II.B. L'anisotropie de fluorescence

### II.B.1. Principe et anisotropie statique

Lors de l'excitation par une source lumineuse polarisée, les fluorophores dont le dipôle est orienté parallèlement au vecteur de la lumière incidente sont excités de manière préférentielle. C'est la photosélection des fluorophores. Si les molécules excitées étaient totalement immobiles, les photons libérés seraient émis suivant un vecteur parallèle au vecteur électrique de la lumière incidente. La fluorescence n'étant pas instantanée, les fluorophores vont subir un mouvement angulaire entre l'excitation et l'émission et l'orientation du vecteur d'émission des photons sera donc modifiée. Ce changement d'orientation conduit à une dépolarisation de la lumière émise (cf. figure 47). L'anisotropie est un paramètre permettant d'accéder à la diffusion rotationnelle des molécules. L'anisotropie ( $r$ ) se calcule avec la formule suivante :

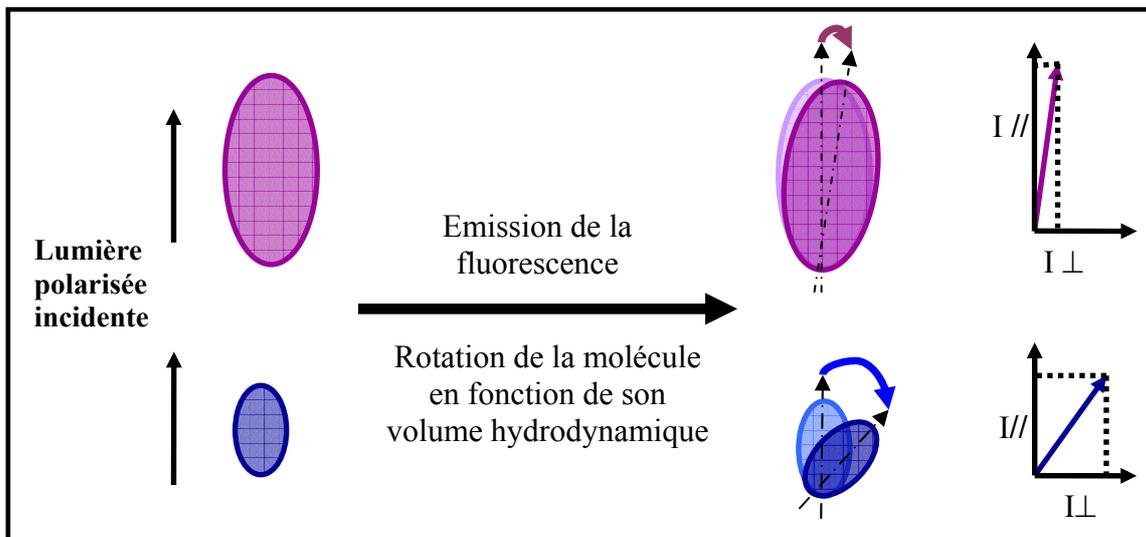
$$r = I_{//} - I_{\perp} / (I_{//} + 2 I_{\perp}) \quad (3)$$

Où  $I_{//}$  est l'intensité de l'émission mesurée parallèlement à l'intensité incidente et  $I_{\perp}$  l'intensité mesurée perpendiculairement. L'anisotropie est un rapport d'intensité, elle est sans dimension et ne dépend pas de la concentration en fluorophore. De plus, l'anisotropie est une valeur additive. Si plusieurs espèces fluorescentes avec des anisotropies différentes sont mélangées en solution, l'anisotropie mesurée est une moyenne donnée par :

$$\langle r \rangle = \sum f_i \times r_i \quad (4)$$

Où  $r_i$  représente l'anisotropie de chacune des espèces et  $f_i$  la fraction qui lui est associée. L'anisotropie de fluorescence est aussi influencée par la viscosité et la température qui agissent sur la diffusion des molécules en solution.

L'anisotropie statique est une valeur moyenne dont chacune des composantes est reliée à un type de mouvement. Ces mouvements peuvent être par exemple la diffusion rotationnelle du complexe fluorescent, le mouvement d'un domaine particulier d'une protéine ou bien le mouvement propre du fluorophore dû à la flexibilité du linker qui le relie au complexe. Du fait que l'anisotropie statique est un paramètre moyenné, elle ne va donc pas permettre de déterminer avec précision la taille du complexe fluorescent. Pour séparer les différentes



*Figure 47 : Principe de l'anisotropie de fluorescence*

composantes de l'anisotropie statique, et donc accéder à la taille du complexe fluorescent, il faut réaliser des mesures d'anisotropie résolue en temps.

### ***II.B.2. L'anisotropie résolue en temps***

Pour accéder à l'anisotropie résolue en temps, il faut exciter les fluorophores avec une lumière polarisée incidente pulsée obtenue par exemple par excitation laser. Dans ce cas, on mesure le déclin des intensités // et  $\perp$  et calcule le déclin de l'anisotropie  $r(t)$ . Pour une sphère, le déclin d'anisotropie est donné par :

$$r(t) = r_0 e^{-t/\theta} \quad (5)$$

Où  $r_0$  est l'anisotropie à  $t = 0$ , et  $\theta$  le temps de corrélation de la sphère. En général, le déclin d'anisotropie n'est pas une monoexponentielle, mais correspond à la somme de plusieurs exponentielles, chacune reliées à un temps de corrélation différent (cf. figure 48). Chaque temps de corrélation est relié à un type de mouvement. Le déclin d'anisotropie de fluorescence est modélisé par l'équation suivante :

$$r(t) = r_0 \sum g_i e^{-t/\theta_i} \quad (6)$$

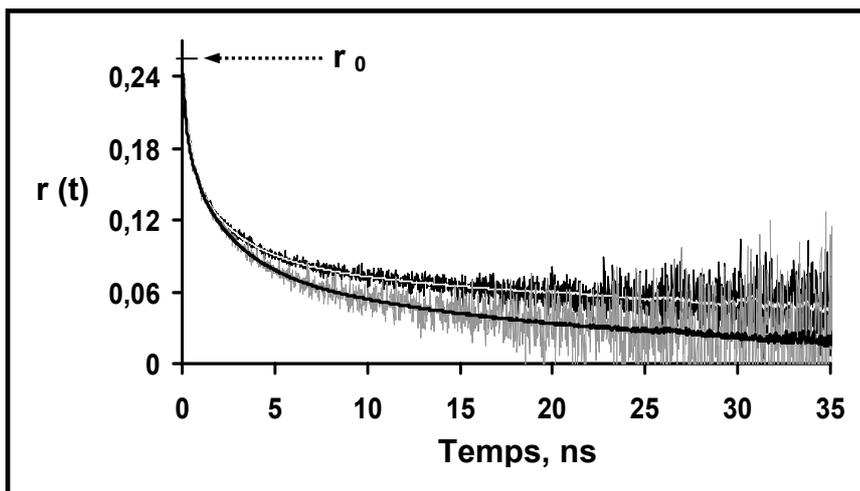
Où  $g_i$  est la fraction de l'amplitude concernée et  $\theta_i$  le temps de corrélation qui lui est associé. La relation entre le temps de corrélation et le volume hydrodynamique est exprimée par la formule d'Einstein :

$$\theta = (\eta V) / (RT) \quad (7)$$

Où  $\eta$  est la viscosité,  $T$  la température en Kelvin,  $R$  la constante des gaz parfaits et  $V$  le volume de l'unité rotationnelle. Lorsque l'observable fluorescente est une protéine globulaire, le temps de corrélation rotationnel peut être approximativement relié à la masse moléculaire ( $M$ ) de la protéine par la relation suivante :

$$\theta = (\eta V) / (RT) = [\eta M(\bar{v} + h)] / (RT) \quad (8)$$

Où  $\bar{v}$  est le volume spécifique de la protéine,  $h$  l'hydratation,  $T$  la température en K,  $R = 8.31 \times 10^7$  en erg/mol $^\circ$ K et la viscosité  $\eta$  est en poise. La valeur de  $h$  pour les protéines est d'environ 0.4 g d'H<sub>2</sub>O par gramme de protéine. L'anisotropie résolue en temps nous permet donc de déterminer avec précision le volume hydrodynamique apparent d'une protéine.



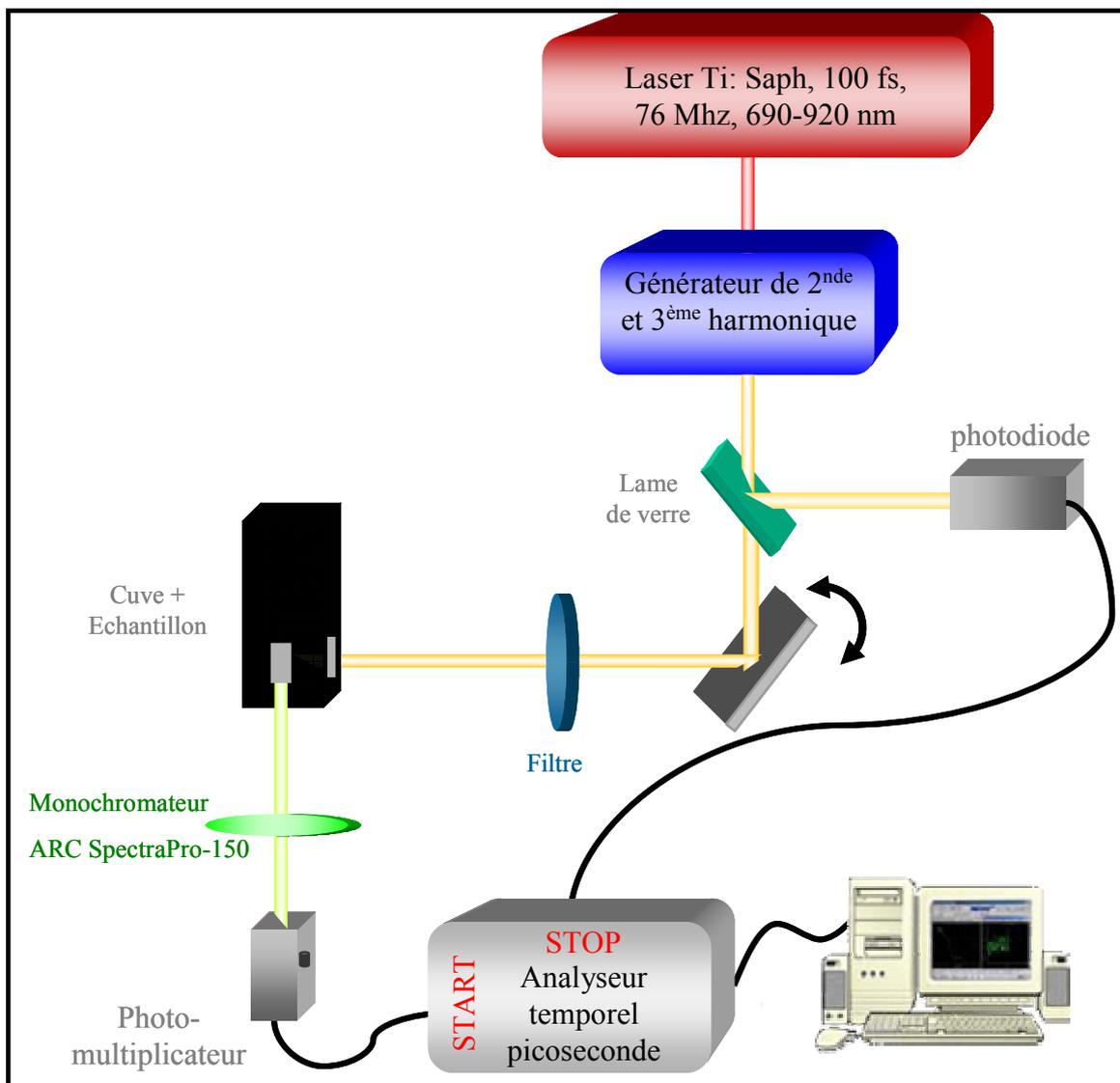
**Figure 48 : Exemple de déclin d'anisotropie**

Dans ce cas, la molécule fluorescente n'est pas une sphère donc l'expression du déclin n'est pas une monoexponentielle, mais la somme de plusieurs exponentielles.

*D'après (Delelis et al., 2008)*

### ***II.B.3. Description du montage utilisé***

La source lumineuse utilisée est un laser Ti :Sapphire femtoseconde (Millenia-pumped Tsunami femtosecond laser, Spectra Physic) avec un taux de répétition de 80 MHz émettant dans l'infrarouge. Le taux de répétition est abaissé à 8MHz par un pulse-picker. Le faisceau laser est ensuite dirigé dans un générateur de seconde ou troisième harmonique qui permet d'obtenir une longueur d'onde d'excitation de 490nm (pour les mesures réalisées en présence de fluorescéine) et 295 nm (pour les mesures réalisées sur l'intégrase seule, où l'on utilise les propriétés de fluorescence du tryptophane). Lorsque cela est nécessaire, des filtres de densités neutres permettent d'atténuer l'intensité de l'excitation. L'émission de la fluorescence est détectée à travers un monochromateur (ARC SpectraPro-150), réglé à 530 nm ou 340 nm pour la fluorescence respectivement de la fluorescéine et du tryptophane (avec une ouverture de 15 nm), par un photomultiplicateur à microcanaux (Hamamatsu modèle R1564U-06) lui-même connecté à un amplificateur (Phillips Scientific modèle 6954, avec un gain de 50). Le pulse de lumière excitatrice est couplé aux mesures grâce à une photodiode à avalanche (Hamamatsu, modèle S4753) (cf. figure 49 pour le montage). L'échelle de temps utilisée pour le comptage des photons émis peut varier, nous avons utilisé échelle de 19,5 picosecondes par canal avec 4096 canaux. Les deux composantes de l'intensité sont mesurées alternativement pendant 30 secondes chacune. Le nombre de cycles varie en fonction de l'intensité de la fluorescence émise. Il est choisi de manière à avoir environ 15 millions de photons pour la composante  $I_{//}$ . La correction de la transmission différentielle des composantes verticales et horizontales à travers le monochromateur ( $G_{\text{facteur}} = I_{VV} / I_{VH}$ ) est déterminée à partir d'une solution de concentration connue de fluorescéine ou de N-acétyl-tryptophanamide. Les microcuves sont thermostatées par un bain circulant (Haake, type F3). Les paramètres du déclin d'anisotropie sont déduits des paramètres des déclins des intensités parallèles  $I_{VV}(t)$  et perpendiculaires  $I_{VH}(t)$ . La distribution des différents temps de corrélation rotationnels est calculée à partir de ces données par la méthode du maximum d'entropie (Brochon, 1994). Le volume hydrodynamique des molécules est ensuite déduit de l'équation 8.



*Figure 49 : Montage utilisé pour les mesures de TFA*

## II.C. Le Transfert d'Énergie Résonant de Fluorescence (FRET)

Ce processus intervient lorsque le spectre d'émission d'un fluorophore, appelé le donneur, possède un recouvrement avec le spectre d'absorption d'une autre molécule, appelée l'accepteur et que les deux molécules sont proches dans l'espace. L'accepteur n'est pas obligatoirement un fluorophore. Le transfert d'énergie résonant n'implique pas une émission de photon par l'accepteur. Dans ce processus, le transfert d'énergie n'est pas réalisé par l'intermédiaire d'un photon, mais par couplage de l'accepteur et du donneur par des interactions dipôle-dipôle. Ce couplage, de part sa nature non radiative, implique aussi une proximité spatiale entre le donneur et l'accepteur (typiquement  $< 70 \text{ \AA}$ ).

Le taux de FRET est donné par :

$$k_T(r) = (1/\tau_D) \times (R_0/r)^6 \quad (9)$$

Où  $\tau_D$  est le temps de vie du donneur en l'absence de FRET,  $R_0$  est la distance de Förster et  $r$  la distance entre l'accepteur et le donneur. La distance de Förster correspond à la distance pour un couple donneur / accepteur pour laquelle l'efficacité de FRET est de 50%.

L'efficacité du FRET correspond à la fraction de photon absorbée par le donneur et transférée à l'accepteur. Elle s'exprime sous la forme d'un ratio entre le taux de FRET et le taux de déclin total du donneur. Ce dernier est influencé par le taux de FRET et les taux de déclins radiatif et non radiatif intrinsèques au donneur. L'efficacité est donnée par :

$$E = k_T(r) / (\tau_D^{-1} + k_T(r)) = R_0^6 / (R_0^6 + r^6) \quad (10)$$

Où  $k_T(r)$  est le taux de FRET.

L'efficacité du FRET est souvent mesurée en utilisant les intensités relatives de fluorescence du donneur en absence ( $F_D$ ) et en présence de l'accepteur ( $F_{DA}$ ) :

$$E = 1 - (F_{DA} / F_D) \quad (11)$$

Nous avons utilisé cette expression lors de nos mesures de FRET entre la Fluorescéine couplée à l'ADN et le TAMRA couplé à l'IN. La distance de Förster de ce couple est d'environ  $51 \text{ \AA}$ . Les intensités de fluorescence du donneur en absence et en présence de l'accepteur ont été mesurées sur le Carry Eclipse. Le protocole est détaillé dans le matériel et

méthode de l'article présenté dans la partie 2 des résultats : « Etude du mécanisme de reconnaissance IN / ADN : un mode d'interaction à l'ADN spécifique mis en évidence par une IN marquée avec une technique enzymatique ».

## II.D. La Spectroscopie de Fluorescence Corrélée (FCS)

Dans le montage mis au point au laboratoire, qui permet de réaliser des mesures de FCS, les fluorophores sont excités selon un mode biphotonique. Dans cette partie, nous allons présenter brièvement la technique de la FCS.

### II.D.1. Fluorescence sous excitation biphotonique

Les fluorophores sont généralement excités lors de l'absorption d'un photon qui possède une longueur d'onde comprise dans la bande d'absorption du fluorophore. Lorsque l'excitation est menée sous forme pulsée avec des lasers puissants et à des fréquences de l'ordre de la femtoseconde, les fluorophores peuvent être alors excités par l'absorption quasi-simultanée de deux photons de longueur d'onde plus grande (d'énergie plus basse) pour atteindre le même état d'excitation (cf. figure 50 A). L'énergie fournie par les deux (ou plus) photons doit être équivalente à celle fournie lors de l'excitation par un seul photon.

La probabilité d'excitation par absorption de deux photons  $\sigma_{\text{abs2p}}$ , et donc celle de l'émission résultante, est quantifiée par la section efficace d'absorption à deux photons. Le nombre de photons absorbés par molécules et par unité de temps est proportionnel au carré de l'intensité instantané. Il est donné par la formule suivante :

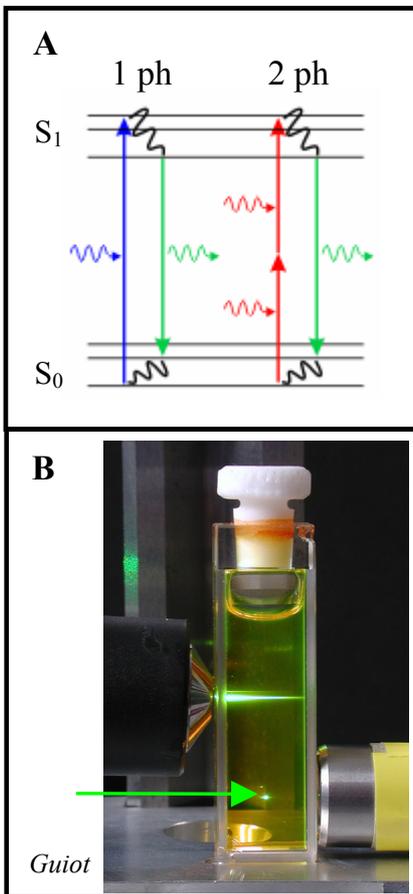
$$N_{\text{abs2p}}(t) = \sigma_{\text{abs2p}} \times I^2(t) \quad (12)$$

Avec  $I(t)$  l'intensité instantanée. La dépendance quadratique à l'intensité vient du fait que les deux photons doivent être quasi-simultanément absorbés pour induire la transition. La fluorescence émise par la molécule, par unité de temps est alors de :

$$F(t) = \frac{1}{2} \eta N_{\text{abs2p}}(t) = \frac{1}{2} \eta \sigma_{\text{abs2p}} \times I^2(t) \quad (13)$$

Avec  $\eta$  le rendement quantique de fluorescence de l'état excité et le facteur  $\frac{1}{2}$  traduit le fait que deux photons sont absorbés pour chaque photon émis. Étant données les très faibles probabilités de transition à deux photons, le flux de photons incident doit être très élevé afin d'obtenir une bonne efficacité d'excitation de la fluorescence.

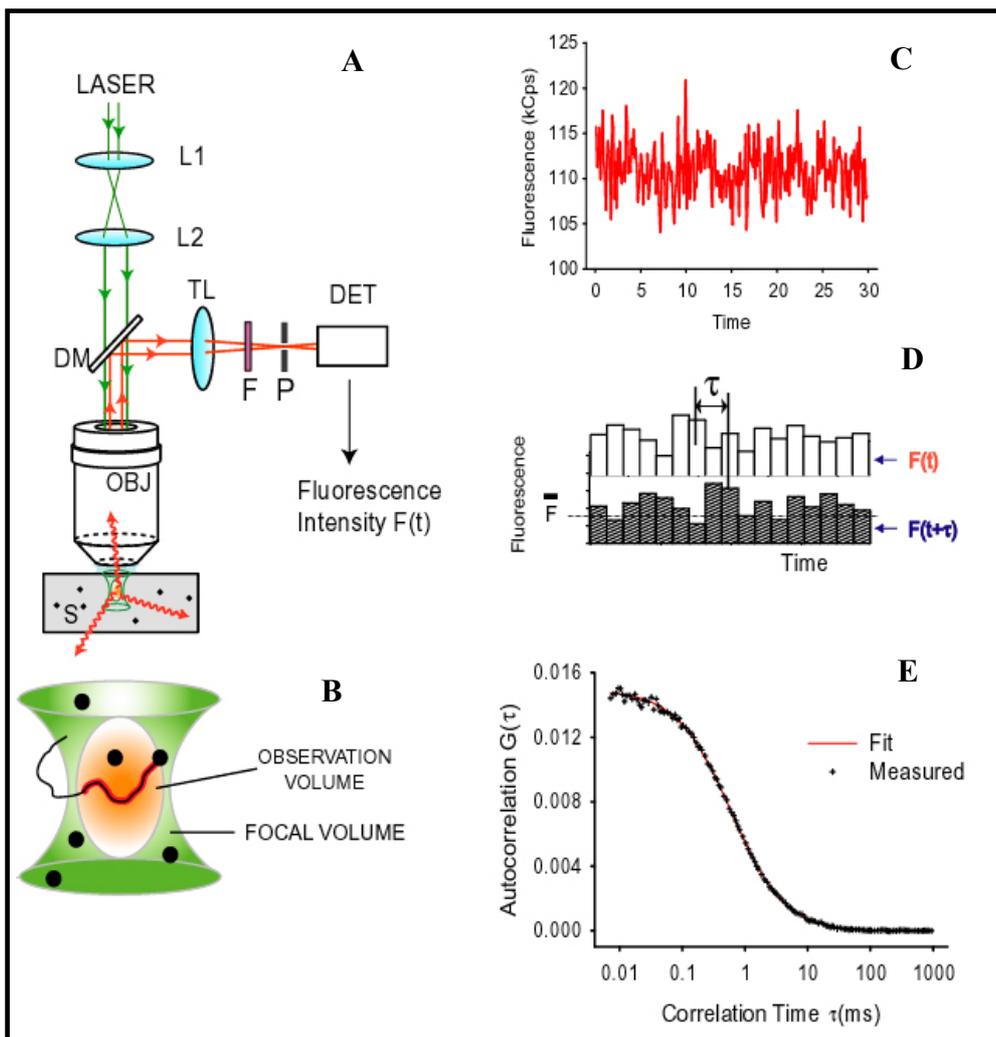
La résolution spatiale de l'intensité d'excitation et de la détection est décrite par la PSF (Point-Spread Function) qui correspond à l'image d'un point à travers un microscope.



**Figure 50 : La fluorescence sous excitation bi-photonique**

**A** : Diagramme de Jablonski comparé entre l'excitation mono et biphotonique. Sous excitation biphotonique, l'énergie des photons excitateurs est plus basse.

**B** : Comparaison du confinement de l'excitation sous excitation biphotonique (flèche verte).



**Figure 51 : Principe de la FCS**

**A** : Schéma théorique du montage de mesure de FCS

**B** : Schéma du volume d'excitation.

**C** : Courbe de mesure de la variation de l'intensité de fluorescence.

**D** : Comparaison de la courbe de la variation de l'intensité en fonction de  $t$  et  $t+\tau$ .

**E** : Courbe d'auto-corrélation résultante de la corrélation de  $F(t)$  et de  $F(t+\tau)$ .

Sous une excitation biphotonique, la PSF effective dépend seulement de la distribution spatiale de l'intensité d'excitation :

$$I_{\text{exc}}^2(x,y,z) = I_0^2 \text{PSF}_{2p}(x,y,z) \quad (14)$$

Avec  $I_0$  l'intensité maximale au centre du volume focal.

La PSF dépend également de la longueur d'onde d'excitation  $\lambda$  et de l'ouverture numérique de l'objectif du microscope utilisé (ON) pour l'observation. De ce fait, en connaissant ces données, on peut approximer la résolution spatiale de la PSF. Des expressions approchées de la largeur de la PSF (donnée qui correspond à la largeur à mi-hauteur ou largeur à 1/e) peuvent être déterminées en réalisant une modélisation du faisceau laser excitateur par un faisceau gaussien dans les directions axiales et latérales :

$$\text{PSF}_r = \omega_0 \approx \lambda / (2 \text{ON})$$

$$\text{PSF}_z = \omega_z \approx (1.3 n \lambda) / (\text{ON}^2) \quad (15)$$

De ce fait, l'excitation en deux photons présente l'avantage d'avoir une résolution dans l'espace fine (cf. figure 50 B). De ce fait, l'échantillon n'est pas excité en totalité. Cette propriété permet de réaliser des acquisitions sur chaque plan focal et de réaliser des reconstructions en 3 Dimension qui présente un intérêt majeur en biologie cellulaire.

### ***II.D.2. Principe de la Spectroscopie de Fluorescence Corrélée (FCS)***

La FCS est basée sur la mesure de la fluctuation de la fluorescence temporaire observée dans un petit volume d'observation. Des volumes d'observation de l'ordre du femtolitre sont obtenus avec une excitation localisée par un système d'excitation multiphotonique (cf. figure 51 A et B). Des burst de photons sont observés comme étant associés à un fluorophore qui diffuse au travers du volume d'excitation. Cette méthode associée à un système de détection très performant est extrêmement sensible ; quelques fluorophores seulement peuvent être observés à un temps donné. La FCS étant une spectroscopie de fluctuation, les solutions ne doivent contenir que de faibles concentrations de fluorophores de manière à avoir un nombre réduit de fluorophores présents simultanément dans le volume d'excitation.

L'intensité est mesurée en fonction du temps ( $I(t)$ ), cette fonction est comparée avec l'intensité observée à un temps  $\tau$  très court plus tard ( $I(t+\tau)$ ) (cf. figure 51 D et E). La fonction d'autocorrélation résultante de cette comparaison est calculée selon l'équation :

$$g(\tau) = \langle I(t) \times I(t+\tau) \rangle / \langle I(t) \rangle^2 \quad (16)$$

Où  $I(t)$  est le nombre de photons émis par fluorescence détectés par unité de temps. En considérant que la distribution de l'intensité d'excitation suit une courbe gaussienne dans les trois dimensions, la fonction de FCS, pour un mouvement de diffusion libre de type brownien est donnée par l'équation :

$$g(\tau) = 1/N \times (1 + \tau/\tau_D)^{-1} \times (1 + \omega_0^2/z_0^2 \times \tau/\tau_D)^{-1/2} \quad (17)$$

Où  $N$  est le nombre moyen de fluorophores dans le volume d'excitation,  $\tau_D$  est le temps de diffusion translationnel.  $\omega_0$  et  $z_0$  sont respectivement les dimensions latérales et axiales du volume d'excitation.

$\tau_D$  est déterminé à partir des courbes d'autocorrélation en fittant ces données par la fonction 17 avec un algorithme non linéaire de Levenberg-Marquardt. Etant donné que notre montage est sous excitation biphotonique, le coefficient de diffusion  $D$  est calculé selon l'équation :

$$\tau_D = \omega_0^2 / 8D \quad (18)$$

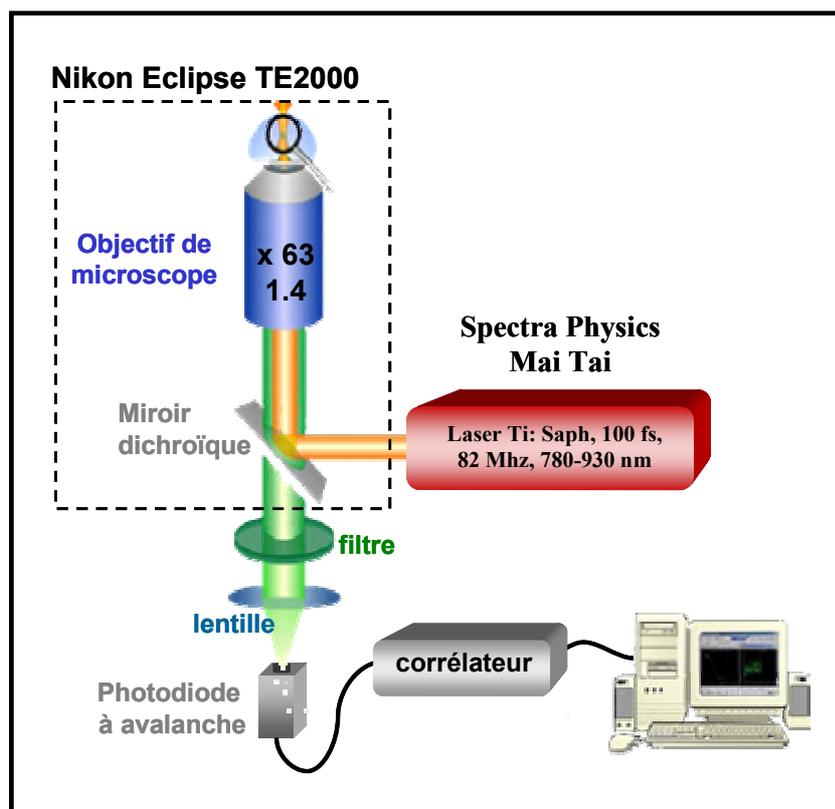
Lorsque  $\tau = 0$ , l'équation 17 donne

$$g(0) = 1 / N. \quad (18)$$

L'ordonnée à l'origine de la fonction d'autocorrélation  $g(\tau)$  permet de déduire le nombre moyen de fluorophores présents dans le volume d'excitation. Lorsque l'on connaît le volume d'excitation, il est possible d'approximer la concentration de la solution observée en fluorophores.

Le coefficient de diffusion  $D$  est relié à la taille des molécules fluorescentes observées par la relation suivante:

$$D = kT / 6\pi\eta R \quad (21)$$



*Figure 52 : Montage utilisé pour les mesures de FCS*

Le montage est réalisé afin d'avoir une excitation en mode biphotonique.

Où  $k$  est la constante de Boltzmann,  $T$  est la température,  $\eta$  est la viscosité du solvant et  $R$  le rayon hydrodynamique.  $R$  est relié au poids moléculaire et au volume spécifique  $\bar{v}$  par la relation suivante :

$$R = ((3MW \bar{v}) / 4\pi)^{1/3} \quad (22)$$

Par exemple, lorsque la taille du complexe fluorescent double, le coefficient de diffusion augmente d'un facteur :  $(2)^{1/3} = 1,26$ .

### ***II.D.3. Description du montage utilisé***

Les mesures de FCS sont réalisées sous une excitation biphotonique, avec un système comprenant un laser Ti :Sapphire laser (pulse : 100fs) à 80 Mhz couplé avec une pompe Millennia (Spectra physics) et un microscope inversé Nikon TE2000 (cf. figure 52).

Avant l'entrée au travers du port de l'épifluorescence du microscope, le rayon laser est élargi avec par un système afocal de deux lentilles qui se trouve derrière l'objectif (Nikon, Plan Apo, 100×, ouverture numérique de 1,4, immersion à huile). Les réglages sont optimisés de telle sorte d'avoir une diffraction minimale au point focal. Les mesures sont faites dans un volume de 50µL de solution déposée sur une lamelle. La fluorescence émise est collectée par le même objectif et est séparée de la lumière d'émission par un miroir dichroïque (Chroma 700DCSPXR) et focalisé sur une photodiode à avalanche (SPCM-AQR-14, Perkin Elmer). Un deuxième filtre est ajouté afin d'éliminer les photons résiduels potentiellement contaminants de la lumière d'excitation (Chroma 2P-Emitter E700SP). Le détecteur est connecté à un corrélateur digital (ALV 6000) qui calcule la fonction d'autocorrélation normalisée  $g(\tau)$  reliée aux fluctuations de la fluorescence selon l'équation 16.

La calibration du volume d'excitation a été réalisée avec une solution de TAMRA à 10 nM dilué dans de l'eau ( $D = 2,8 \times 10^{-10} \text{ m}^2\text{s}^{-1}$ ). Les dimensions du volume d'excitation  $\omega_0$  et  $z_0$  ont été estimées respectivement à 0,45 et 1,5 µm. La puissance d'excitation est ajustée en utilisant des filtres de densité neutres. Concernant ce montage, une puissance de 10mW a été déterminée comme suffisante pour réaliser l'excitation à deux photons du TAMRA seul ou lié à l'IN de PFV-1. Les enregistrements sont de 5 min, avec une moyenne de 10 cycles de 30 secondes par acquisition.



# **RESULTATS**

Mes travaux de thèse se sont déroulés selon deux axes. Le premier est basé sur l'étude des paramètres structuraux et cinétiques qui régissent la réaction de 3'-processing catalysée par les IN rétrovirales. Cet axe d'étude a donné lieu à la publication de deux articles dans la revue « Journal of Biological Chemistry » en août 2006 (*2006 Aug 11;281(32):22707-19*) et août 2008 (*2008 Aug 12. [Epub ahead of print]*). Le second axe est basé sur l'implication du domaine N-ter de l'IN du VIH-1 dans l'assemblage des multimères IN / ADN viral catalytiquement actifs. Cette étude a été menée dans les deux contextes cationiques permissifs pour le 3'-processing de l'IN. Les résultats obtenus dans le cadre de cette étude seront présentés sous la forme d'un manuscrit qui est en fin de rédaction.

Chacun des trois manuscrits présentés sera précédé d'une introduction qui replace l'article dans son contexte scientifique et sera suivi d'une discussion relative aux résultats obtenus. Une discussion plus générale suit cette partie.

# ***I. Relation entre l'état oligomérique de l'IN du VIH-1 sur l'ADN et son activité enzymatique***

## **I.A. Introduction**

La réaction de 3'-processing des deux extrémités de l'ADN viral est la première étape catalysée par l'IN du VIH-1. Cette réaction de clivage endonucléolytique aboutit à la formation de deux produits : l'ADN viral processé et le dinucléotide GT. Ce dernier est libéré dans le milieu réactionnel, alors que l'ADN processé devient le substrat pour la deuxième réaction catalysée par l'IN : le transfert de brin. Ces deux réactions peuvent être reproduites *in vitro* avec des ADN courts mimant l'extrémité LTR de l'ADN viral.

Le déterminisme du mécanisme réactionnel de l'étape de 3'-processing a fait l'objet d'un grand nombre d'études. Cependant, les déterminants structuraux de l'activité catalytique de l'IN restent encore sujets à controverse, notamment sur ses états multimériques actifs. En effet, les structures cristallographiques des doubles domaines de l'IN du VIH-1 (IN<sup>1-212</sup> et IN<sup>55-288</sup>) ainsi que la structure de domaine de l'IN en présence d'interactants cellulaires comme LEDGF/p75 montrent une structure dimérique conservée pour le CC. Au niveau de ce dimère, les deux sites actifs sont séparés d'environ 35 Å. Lors de l'intégration concertée, l'IN intègre les deux extrémités de l'ADN viral avec un écart constant de 5 paires de bases, qui correspondent à une distance d'environ 15 Å. Ces données associées à des tests d'activité d'intégration concertée *in vitro* avec des multimères d'IN cross-linkés, montrent que pour cette étape, l'état multimérique actif de l'IN du VIH-1 est une forme de taille supérieure ou égale à un tétramère d'IN. Un état multimérique aussi important est-il également nécessaire pour la catalyse du 3'-processing ? De plus, la question se pose quant à l'influence des propriétés auto-associatives de l'IN sur l'efficacité de son activité enzymatique. Actuellement la plupart des tests enzymatiques effectués *in vitro* sont réalisés avec un excès d'IN par rapport à son substrat ADN. Etant donné la faible solubilité de cette enzyme, nous avons voulu savoir dans quelle mesure les multimères de haut poids ou agrégats, dont la présence est favorisée à ces fortes concentrations, pouvaient influencer l'activité catalytique mesurée.

Dans cet article, nous présentons un nouveau test d'activité permettant d'observer sur un même échantillon la fixation à l'ADN et l'activité catalytique de 3'-processing résultante. Ce test présente l'avantage de pouvoir séparer l'analyse des deux événements. Ce test permet également de réaliser des cinétiques en temps réel de l'activité 3'-processing sur un même

échantillon et d'obtenir des paramètres catalytiques. Il est basé sur les propriétés de l'anisotropie de fluorescence ( $r$ ) qui est une mesure utilisant les propriétés de diffusion rotationnelle des molécules fluorescentes. Le principe du test est le suivant : un oligonucléotide mimant l'extrémité LTR de l'ADN viral est marqué par une fluorescéine à son extrémité 3' CAGT au niveau du T. Il est alors possible de suivre la fixation de l'IN sur l'ADN par une augmentation de  $r$ , puis la libération du dinucléotide clivé par une diminution de  $r$  après passage à une température permissive pour l'activité (37°C). Les différences de tailles entre les différentes formes fluorescentes, ADN seul, complexe IN / ADN et dinucléotide, sont telles que les différences d'anisotropies permettent d'observer dans une première étape la fixation de l'IN sur l'ADN et d'évaluer la fraction de complexe formé selon les ratio IN / ADN mis en présence ; dans un deuxième temps, de calculer la fraction de dinucléotide GT fluorescent libéré en fonction du temps et d'en déduire l'activité catalytique de l'IN. Le paramètre statique  $r$  donne aussi une indication sur la taille des complexes IN / ADN présents dans la solution. Une détermination plus précise de leur taille a été réalisée à partir de mesures d'anisotropie de fluorescence résolue en temps sur les différents échantillons utilisés lors de ce test.

## **I.B. Manuscrit**

Ce manuscrit a été publié dans « the Journal of Biological Chemistry » vol 281, n°32, pp. 22707-22719, le 11 août 2006.

## Relationship between the Oligomeric Status of HIV-1 Integrase on DNA and Enzymatic Activity\*

Received for publication, March 8, 2006, and in revised form, May 24, 2006. Published, JBC Papers in Press, June 13, 2006, DOI 10.1074/jbc.M602198200

Elvire Guiot<sup>‡</sup>, Kevin Carayon<sup>‡</sup>, Olivier Delelis<sup>‡</sup>, Françoise Simon<sup>‡</sup>, Patrick Tauc<sup>‡</sup>, Evgenii Zubin<sup>§</sup>, Marina Gottikh<sup>§</sup>, Jean-François Mouscadet<sup>‡</sup>, Jean-Claude Brochon<sup>‡</sup>, and Eric Deprez<sup>‡1</sup>

From the <sup>‡</sup>Laboratoire de Biotechnologie et Pharmacologie Génétique Appliquée, CNRS, UMR8113, Ecole Normale Supérieure de Cachan, 61 av du Président Wilson, 94235 Cachan, France and the <sup>§</sup>Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119992 Moscow, Russia

The 3'-processing of the extremities of viral DNA is the first of two reactions catalyzed by HIV-1 integrase (IN). High order IN multimers (tetramers) are required for complete integration, but it remains unclear which oligomer is responsible for the 3'-processing reaction. Moreover, IN tends to aggregate, and it is unknown whether the polymerization or aggregation of this enzyme on DNA is detrimental or beneficial for activity. We have developed a fluorescence assay based on anisotropy for monitoring release of the terminal dinucleotide product in real-time. Because the initial anisotropy value obtained after DNA binding and before catalysis depends on the fractional saturation of DNA sites and the size of IN:DNA complexes, this approach can be used to study the relationship between activity and binding/multimerization parameters in the same assay. By increasing the IN:DNA ratio, we found that the anisotropy increased but the 3'-processing activity displayed a characteristic bell-shaped behavior. The anisotropy values obtained in the first phase were predictive of subsequent activity and accounted for the number of complexes. Interestingly, activity peaked and then decreased in the second phase, whereas anisotropy continued to increase. Time-resolved fluorescence anisotropy studies showed that the most competent form for catalysis corresponds to a dimer bound to one viral DNA end, whereas higher order complexes such as aggregates predominate during the second phase when activity drops off. We conclude that a single IN dimer at each extremity of viral DNA molecules is required for 3'-processing, with a dimer of dimers responsible for the subsequent full integration.

The integration of a DNA copy of the HIV-1<sup>2</sup> genome into the host genome is a crucial step in the life cycle of the retrovirus. Integrase (IN) is responsible for the two consecutive reac-

tions that constitute the overall integration process. The first of these two reactions is 3'-processing, which involves cleavage of the 3'-terminal GT dinucleotide at each extremity of the viral DNA. The hydroxyl groups of newly recessed 3'-ends are then used in the second reaction, strand transfer, for the covalent joining of viral and target DNAs, resulting in full-site integration. IN is sufficient for catalysis of the 3'-processing reaction *in vitro*, using short-length oligodeoxynucleotides (ODNs) that mimic one viral long terminal repeat (LTR) in the presence of the metallic cofactor Mg<sup>2+</sup>. This reaction generates two products: the viral DNA containing the recessed extremity and the GT dinucleotide. One of the two products, the processed viral DNA, as well as the target DNA serve as substrates for the subsequent joining reaction.

IN belongs to the superfamily of polynucleotidyl transferases. Its catalytic core domain contains a triad of acidic residues constituting the D,D-35-E motif, which is strictly required for catalysis. The catalytic core establishes specific contacts with the viral DNA and, together with the C-terminal domain, is involved in DNA binding (1–4). The 3'-processing reaction is highly specific, and the terminal 13 bp of the LTR play a key role for Mg<sup>2+</sup>-dependent 3'-processing in terms of reaction specificity (4, 5). In particular, the CA sequence preceding the GT dinucleotide cleaved by IN is strictly required. Reaction specificity seems to depend on the catalytic step, because no significant difference in affinity is observed *in vitro* for different DNA sequences (2, 3, 5, 6). The catalytic mechanism of IN has been extensively studied, but the structural determinants of IN activity remain unclear, and controversy remains concerning the multimeric status of active IN. Single- and double-domain crystallographic structures have been produced for IN, alone or complexed with the IN-binding domain of the cellular partner lens epithelium-derived growth factor, and all these structures display a conserved dimeric structure of the catalytic core (7–11). Based on topological considerations concerning the distance separating the active sites of the two protomers, which is more than double the distance separating the two joining sites on the target DNA (separated by 5 bp), a multimeric state of an order higher than dimers, tetramers at least, is required for complete integration. However, it remains unclear whether such a higher order multimeric state is also required for 3'-processing activity and how this activity is modulated by the self-association properties of IN. Moreover, most enzymatic studies are carried out with an excess of IN over DNA substrate. Taking into account the low solubility of IN,

\* This work was supported by grants from the TrioH European Project (FP6 Grant 503480), the CNRS, the French National Agency for Research against AIDS, and the Russian Foundation for Basic Research (Grants 04-04-22000 and 05-04-48743). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> To whom correspondence should be addressed. Tel.: 33-147-40-23-94; Fax: 33-147-40-76-84; E-mail: deprez@lbpa.ens-cachan.fr.

<sup>2</sup> The abbreviations used are: HIV, human immunodeficiency virus; IN, integrase; LTR, long terminal repeat; ODN, oligodeoxynucleotide; ds, double-stranded; ss, single-stranded; *r*, steady-state fluorescence anisotropy; TFA, time-resolved fluorescence anisotropy; TAMRA, carboxytetramethylrhodamine; wt, wild-type.

## The IN 3'-Processing Reaction Mechanism

we investigated whether high order multimers or aggregated forms of IN favored by high protein concentrations in solution were detrimental or beneficial for activity.

We developed an assay for monitoring DNA binding and subsequent 3'-processing in the same sample. This assay makes possible the separation of binding and catalytic parameters and the study of real-time kinetics. It is based on steady-state fluorescence anisotropy ( $r$ ), which is sensitive to rotational diffusion, and thus suitable for studies aiming to identify structural modifications leading to a significant change in molecular size (12–15). Using a fluorescent probe covalently linked to the GT dinucleotide, this makes it possible to follow DNA binding and dinucleotide release, because both steps strongly influence molecular size of the fluorescent moiety. The anisotropy-based technique is also highly suitable for studies of the relationship between the overall size of IN·DNA complexes and activity. We found that, for low IN:DNA ratios, the  $r$  values obtained after DNA binding and before catalysis were fully predictive of subsequent IN activity, according to the fractional saturation function. For high IN:DNA ratios, anisotropy continued to increase, but 3'-processing activity decreased. Because  $r$  depends on both fractional saturation and the molecular size of complexes at saturation of DNA sites, our results show that high order multimers or aggregated states of IN are detrimental to 3'-processing activity. Activity levels were highest for non-aggregative smaller species. A more precise characterization of catalytically competent complexes by time-resolved fluorescence anisotropy (TFA) (16–20), confirmed steady-state data and led to the identification of dimeric forms as the most active forms for 3'-processing. Our results also highlight that only the IN:DNA ratio but not the IN concentration *per se* determines the aggregation properties and thus the activity of IN and that DNA binding stimulates the self-organization of IN to give a catalytically competent non-aggregative form.

### EXPERIMENTAL PROCEDURES

**Oligonucleotides and Nomenclature**—Unlabeled and fluorescein-labeled single-stranded (ss) ODNs were purchased from Eurogentec (Liege, Belgium) (except for the 2'-aminouridine-containing ODN) and purified by electrophoresis in acrylamide gels. The specific HIV sequence was 5'-GTGTGGAAAATCTCTAGCAGT-3' (the bases removed by IN are underlined). This sequence was denoted *a*, and the complementary non-processed strand was denoted *b*. Fluorescein (F) was attached at the 5'- or the 3'-end of strand *a* or *b*, via a 6-carbon linker. As an example, the specific double-stranded (ds) ODN used to monitor activity was called HIV-*a*3F, indicating that fluorescein was attached to the 3'-end of strand *a*. HIV-GTGT-*a*3F was identical except that the four terminal bases at the 3'-end of strand *a* were GTGT rather than the canonical sequence CAGT. The nonspecific sequence, NS-TTCC, was 5'-ACCTATGCGCCGCTAGATTCC-3' (strand *a*). Two other nonspecific ODNs with different 3'-ends of strand *a* were derived: NS-CACC and NS-CAGT. ds ODNs were obtained by mixing equimolar amounts of complementary strands in 20 mM Tris-HCl (pH 7.2), 100 mM NaCl. The mixture was heated to 85 °C for 5 min, and annealing was allowed by slow cooling to 25 °C.

The uridine-containing ODN (corresponding to strand *a* of the DNA substrate HIV-*a*UF) contains a fluorescein attached to the 2'-amino group of the 3'-terminal 2'-aminouridine. It was synthesized as follows: the solid support was prepared as previously described (21). Briefly, 2'-deoxy-5'-*O*-4,4'-dimethoxytrityl-2'-*O*-trifluoroacetamidouridine (0.2 mmol) was co-evaporated with pyridine (3 × 5 ml) and dissolved in dry pyridine (2 ml). Succinylated long-chain alkylamine-controlled pore glass (500 Å) (200 mg), 2,4,6-triisopropylbenzenesulfonyl chloride (0.6 mmol), and 1-methylimidazole (1.2 mmol) were added, and the mixture was incubated for 3 h at 25 °C. The support was then filtered and washed successively with pyridine, CH<sub>2</sub>Cl<sub>2</sub>, and ether. Starting from 2'-deoxy-5'-*O*-4,4'-dimethoxytrityl-2'-*O*-trifluoroacetamidouridine-derivatized long-chain alkylamine-controlled pore glass (500 Å), a 21-mer ODN was assembled on an ABI394B DNA synthesizer, by the phosphoramidite method, according to the manufacturer's recommendation. Protected 2'-*O*-deoxyribonucleoside phosphoramidites and *S*-ethylthiotetrazole were purchased from Glen Research. Ammonia was used for cleavage from the support and deprotection overnight at 55 °C. The reaction mixture was then analyzed by reverse-phase high-performance liquid chromatography in ion-pair mode. For fluorescein labeling, 1 mg of fluorescein isothiocyanate dissolved in 20 μl of dimethylformamide was added to a solution of 40 nmol of 2'-amino ODN in 130 μl of sodium carbonate-bicarbonate (1 M, pH 9):water (v/v, 5:8) (22, 23). The mixture was incubated at 25 °C in the dark for 18 h and then loaded onto a NAP5-Sephadex G-25 column (Amersham Biosciences) pre-equilibrated in water. The ODN was eluted with water and analyzed by reversed-phase high-performance liquid chromatography in ion-pair mode.

**IN Purification and Standard Analysis of the 3'-Processing Reaction**—HIV-1 IN (32 kDa) was purified under native conditions, as previously described (24). Standard IN assay, using <sup>32</sup>P-labeled ODN and gel electrophoresis, was carried out as previously described (24). Gels were analyzed on a STORM 840™ PhosphorImager (Amersham Biosciences) and quantified with ImageQuaNT™ 4.1 software. The 3'-processing activity was calculated as follows: activity (%) = 19-mer/(21-mer + 19-mer) × 100.

**Steady-state Fluorescence Anisotropy Assay**—Steady-state anisotropy values were recorded on a Beacon 2000 instrument (PanVera, Madison, WI), in a cell thermostatically held at 25 or 37 °C, for the DNA-binding step and 3'-processing reaction, respectively. Unless otherwise stated, we studied the formation of IN·DNA complexes by incubating fluorescein-labeled ds ODNs with IN in 20 mM Tris (pH 7.2), 1 mM dithiothreitol, 20 mM NaCl, 5 mM MgCl<sub>2</sub> (the sample volume was 200 μl). The fractional saturation function ( $Y$ ) was calculated as follows,

$$Y = \frac{r - r_{\text{ODN}}}{r_{\text{max}} - r_{\text{ODN}}} \times 100 \quad (\text{Eq. 1})$$

where  $r_{\text{max}}$  and  $r_{\text{ODN}}$  are the anisotropies of IN-bound and free ODN, respectively (no significant concomitant change in fluorescence intensity was observed). After the DNA-binding step, the temperature was raised from 25 to 37 °C for monitoring of

the catalytic process. The 3'-processing activity was assessed by quantifying the decrease in  $r$ . Two independent methods were used for quantification as follows. (i) In fixed-time experiments, the reaction was stopped by adding SDS (0.25% final), disrupting all the IN·DNA complexes in the sample. In such experiments, the solution contained two fluorescent species: the non-processed ODN and the fluorescein-labeled dinucleotide released by the cleavage reaction. The fraction of dinucleotides released ( $F_{\text{dinu}} = [\text{GT}]/[\text{DNA}]_{\text{total}}$ ) is given by Equation 2,

$$F_{\text{dinu}} = \frac{r_{\text{NP}} - r}{r_{\text{NP}} - r_{\text{dinu}}} \quad (\text{Eq. 2})$$

where  $r_{\text{NP}}$  and  $r_{\text{dinu}}$  are the anisotropy values for pure solutions of non-processed ds ODN and dinucleotide, respectively (fluorescence did not change significantly during the reaction). We used the 5'-GT-3'F dinucleotide (Eurogentec) to determine  $r_{\text{dinu}}$ . (ii) In real-time conditions, an additional fluorescent population corresponding to IN complexed with the unprocessed ds ODN, is present in the sample. In this case,  $F_{\text{dinu}}$  was calculated as follows,

$$F_{\text{dinu}} = \frac{r_{t=0} - r}{r_{\text{max}} - r_{\text{dinu}}} \quad (\text{Eq. 3})$$

where  $r_{\text{max}}$  is the characteristic  $r$  value obtained for optimal activity, and  $r_{t=0}$  is the  $r$  value obtained at the end of the DNA-binding step (before the start of the reaction). The 3'-processing activity obtained with Equations 2 and 3 is referred to as Activity<sub>SDS</sub> and Activity<sub>real-time</sub>, respectively. Activity<sub>real-time</sub> was not used if  $r_{t=0}$  was higher than 0.22 (aggregation of IN on DNA not negligible). We analyzed single-turnover kinetics using the Equations 4 and 5,

$$\ln(1 - F_{\text{dinu}}) = -k_{\text{obs}} \times t \quad (\text{Eq. 4})$$

with

$$k_{\text{obs}} = k_{\text{chemistry}} / ((K_{d,\text{app}} / [\text{IN}]_0) + 1) \quad (\text{Eq. 5})$$

where  $k_{\text{chemistry}}$  is the single-turnover rate constant, and  $K_{d,\text{app}}$  is the apparent  $K_d$  (25). Reactions were conducted in the presence of  $\text{Mg}^{2+}$  (not  $\text{Mn}^{2+}$ ) to limit nonspecific hydrolysis products (cleavages at positions -3, -4, etc.), because anisotropy cannot discriminate between these small products (which are minor products with  $\text{Mg}^{2+}$  but not with  $\text{Mn}^{2+}$ ) and the specific GT product. The  $r$  values for the TAMRA-labeled DNA substrate ( $\lambda_{\text{ex}} = 562$  nm and  $\lambda_{\text{em}} = 582$  nm) were recorded on a Cary Eclipse spectrofluorometer (Varian, Mulgrave, Australia) in polarization mode and were compared with values obtained with the same instrument and HIV-*a3F* as the substrate.

$K_{d,\text{app}}$  as a function of ODN length were determined by competition experiments. Fluorescein-labeled ds ODN HIV-*a5F* (4 nm) was preincubated with various concentrations of unlabeled ss or ds HIV ODN (from 10- to 45-mers) in 20 mM Tris buffer (pH 7.2) supplemented with 1 mM dithiothreitol, 20 mM NaCl, and 5 mM  $\text{MgCl}_2$ . IN (150 nM final concentration) was then added, and  $r$  values were recorded.  $\Delta r (= r - r_{\text{ODN}})$  was plotted against competitor concentration to determine  $K_{d,\text{app}}$  (concentration of competitor decreasing the initial  $\Delta r$  value by 50%).

**Time-resolved Fluorescence Experiments**—Time-resolved fluorescence parameters (lifetimes and correlation times) were obtained from the two polarized fluorescence decays  $I_{\perp}(t)$  and  $I_{\parallel}(t)$ , using the time-correlated single photon counting technique. The instrumentation setup was essentially similar to those previously described (16, 17), with modifications: the time scaling was 19.5 ps per channel and 4096 channels were used. The excitation light pulse source was a Ti:sapphire laser (Millennia-pumped Tsunami femtosecond laser, Spectra Physics) (repetition rate: 8 MHz) associated with a second harmonic generator tuned to 490 nm. The emission monochromator (ARC SpectraPro-150) was set to 530 nm ( $\Delta\lambda = 15$  nm). The two polarized components were collected alternately over a period of 30 s (total count of  $I_{\parallel}$ : 15,000,000). The reaction mixture contained 20 mM Tris, pH 7.2, 20 mM NaCl, 5 mM  $\text{MgCl}_2$ , 1 mM dithiothreitol. The correlation time ( $\theta$ ) distributions of free ss or ds ODNs were obtained at ODN concentrations of 10 nM or 0.5  $\mu\text{M}$ . All ODNs (from 10- to 45-mer) mimicked the U5-end of the HIV-1 DNA and were fluorescein-labeled at the 5'-end of strand *a*. IN·DNA complexes were analyzed using ds HIV-*a5F* and various IN:DNA ratios, from 40 to 400. We analyzed both decays,  $I_{\perp}(t)$  and  $I_{\parallel}(t)$ , by the maximum entropy method (26). Fluorescence anisotropy decay is described by Equations 6 and 7,

$$r(t) = \sum_{i=1}^n \rho_i \times e^{-t/\theta_i} \quad (\text{Eq. 6})$$

with

$$\sum_{i=1}^n \rho_i = \rho_0 \quad (\text{Eq. 7})$$

where  $\theta_i$  is the individual rotational correlation time, and  $\rho_i$  is the associated amplitude.  $\rho_0$  was found to lie between 0.36 and 0.37. Normalization of  $\theta$  for a given temperature was performed using,

$$\theta = \eta V / kT \quad (\text{Eq. 8})$$

where  $\eta$  is the viscosity,  $V$  is the volume of the rotating unit,  $k$  is the Boltzmann constant, and  $T$  is the temperature (K).

## RESULTS

**Monitoring 3'-Processing Activity by Steady-state Fluorescence Anisotropy**—Fluorescence anisotropy measurements are based on the principle of photoselective excitation of a fluorophore by a polarized light, providing information about rotational motions of the fluorophore or fluorescently labeled molecule between photon absorption and emission. Some events such as overall rotational diffusion or flexibility are major causes of light depolarization. High levels of anisotropy are generally associated with large molecules or complexes characterized by slow rotational diffusion or low flexibility level. In this study, we used an extrinsic fluorophore covalently linked to DNA to monitor the binding of IN to viral DNA substrate and the subsequent 3'-processing reaction, in the same assay. Both DNA binding and 3'-processing would be expected to have a

## The IN 3'-Processing Reaction Mechanism

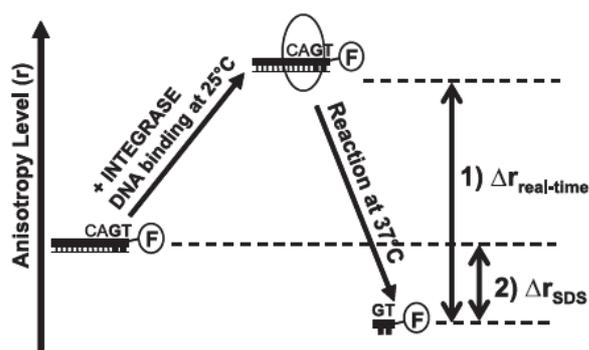


FIGURE 1. Principle of the activity assay based on fluorescence anisotropy. The substrate is a 21-bp HIV ODN labeled with fluorescein at the 3'-end of the CAGT sequence. Upon IN binding to ODN, steady-state anisotropy ( $r$ ) increases. At 37 °C, release of the terminal fluorescein-labeled GT dinucleotide by the 3'-processing reaction leads to a decrease in  $r$ . Activity can then be quantified by two methods: 1) The fraction of released dinucleotides ( $F_{\text{dinu}}$ ) is related to the amplitude  $\Delta r$  and can be estimated from the direct observation of decreases in  $r$  values during the reaction (real-time assay). 2) The reaction can be quenched at any time by adding SDS, which disrupts the IN-ODN complexes (fixed-time assay). In this case,  $F_{\text{dinu}}$  is related to the amplitude  $\Delta r_{\text{SDS}}$ .

significant effect on the anisotropy parameter, because each of these steps has a major effect on the molecular size of the fluorescent moiety. The principle underlying the anisotropy-based assay is summarized in Fig. 1. The binding of IN to a dsODN mimicking one end of the viral DNA increases the steady-state anisotropy value ( $r$ ), most likely by restricting flexibility, allowing the calculation of fractional saturation function. Following DNA binding, the same sample can be shifted to a permissive temperature for 3'-processing activity. When the fluorophore is linked to the 3'-terminal GT dinucleotide, release of the dinucleotide product should significantly decrease  $r$ . The fraction of dinucleotides released can be determined from the real-time decrease in  $r$  observed during the reaction ( $\Delta r_{\text{real-time}}$ , corresponding to the difference between IN-DNA complex and free dinucleotide) or after the disruption of all IN-DNA complexes by the addition of SDS ( $\Delta r_{\text{SDS}}$ , corresponding to the difference between unprocessed dsODN and free dinucleotide). Unlike most standard 3'-processing assays, which are based on quantification of the processed strand (first reaction product), the anisotropy-based assay monitors the released dinucleotide (second reaction product).

We tested 21-mer DNA substrates labeled with fluorescein at one end, in all possible combinations (5'- or 3'-extremity of the  $a$  or  $b$  strand), or labeled on the 2'-amino group of a 3'-terminal 2'-aminouridine (Fig. 2A). The free dsODNs were characterized by  $r$  values of 0.060–0.130 at 25 °C (0.045–0.105 at 37 °C), depending on the location of fluorescein, whereas the fluorescein-labeled dinucleotide was characterized by  $r = 0.02$ . DNA binding on the addition of IN was monitored at 25 °C and led to a significant increase in  $r$  value ( $>0.2$  in the experimental conditions of Fig. 2). Equilibrium was typically reached after 15 min, as previously reported (12, 27). The processing reaction was started by shifting the sample to a temperature of 37 °C, and the initial  $r$  values obtained therefore correspond to  $r_{t=0}$ . In the presence of the divalent cationic cofactor  $\text{Mg}^{2+}$ , all tested ODNs gave similar  $r_{t=0}$  values (0.22), indicating that fluorescein position had no effect on DNA binding by IN (Fig. 2B).

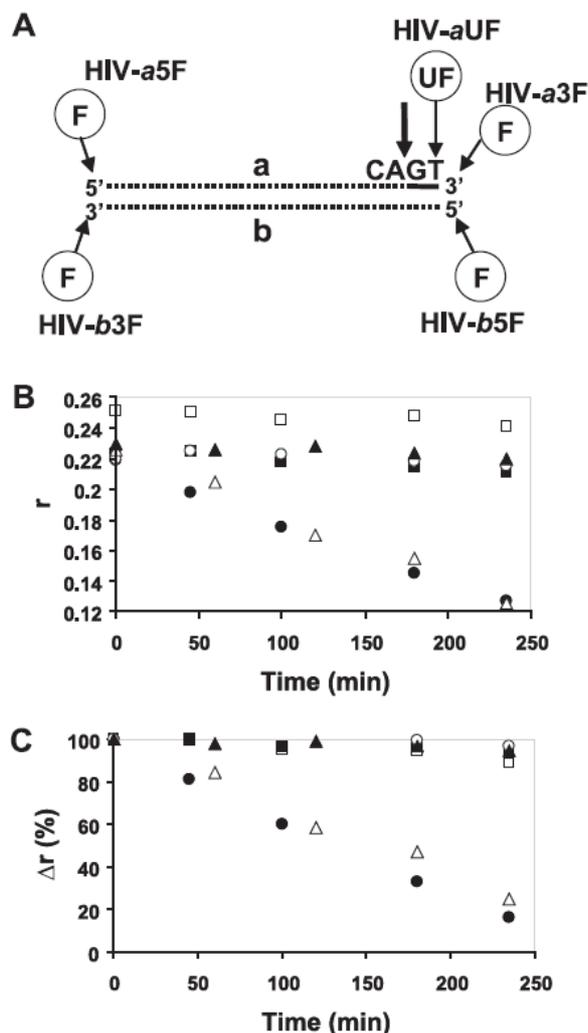


FIGURE 2. The decrease in steady-state fluorescence anisotropy during 3'-processing is dependent on the position of fluorescein on the DNA substrate. A, position of the fluorophore on the various ODNs. Fluorescein was attached to the 3'- or 5'-extremity of each strand ( $a$  or  $b$ , where  $a$  is the strand containing the specific CAGT terminal sequence) giving HIV-a3F, HIV-a5F, HIV-b3F, or HIV-b5F. HIV-aUF has a fluorescein-labeled uridine at the 3'-terminal position of strand  $a$ . The 3'-processing reaction involves the cleavage of the terminal GT dinucleotide, as indicated by the arrow. IN was allowed to bind to the various ODNs for 20 min at 25 °C (3'-processing was undetectable during this preincubation step). The samples were then incubated at 37 °C, and  $r$  values were recorded over 4 h. (B) or  $\Delta r$  (%) (C) was plotted against time.  $\Delta r$  (%) is a relative  $r$  value calculated as follows:  $\Delta r$  (%) =  $(r - r_{\text{ODN}})/(r_{t=0} - r_{\text{ODN}}) \times 100$ . White squares: HIV-a3F, no  $\text{Mg}^{2+}$ ; black circles: HIV-a3F, 5 mM  $\text{Mg}^{2+}$ ; white circles: HIV-b5F, 5 mM  $\text{Mg}^{2+}$ ; black squares: HIV-a5F, 5 mM  $\text{Mg}^{2+}$ ; black triangles: HIV-b3F, 5 mM  $\text{Mg}^{2+}$ ; and white triangles: HIV-aUF, 5 mM  $\text{Mg}^{2+}$ . IN and ODN concentrations were 120 and 4 nM, respectively.

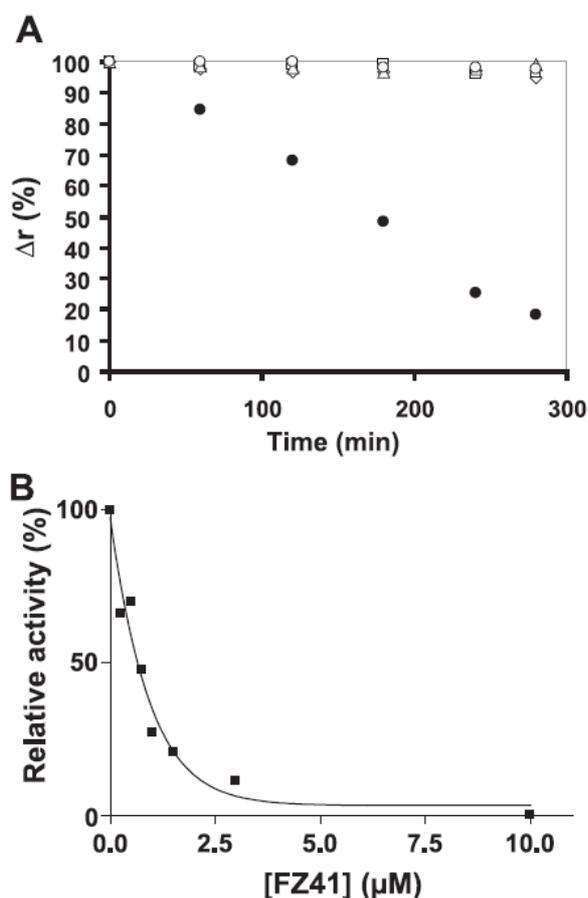
Experiments carried out in the absence of  $\text{Mg}^{2+}$  gave reproducibly larger  $r$  values (0.25). This result confirms that DNA-binding activity of IN is not strictly dependent on  $\text{Mg}^{2+}$ , consistent with other studies (2, 28, 29). Nevertheless, the larger  $r$  value suggests that IN aggregation is favored by an absence of  $\text{Mg}^{2+}$ , as previously reported (17).

The time dependence of  $r$  and  $\Delta r$  under real-time conditions at 37 °C is shown in Fig. 2 (B and C): only the DNA substrates HIV-a3F and HIV-aUF, with the fluorescein directly attached to the small GT reaction product, displayed significant

decreases in  $r$  or  $\Delta r$ , and these decreases were strictly related to the presence of the metallic cofactor. In contrast, HIV- $\alpha 5F$ , HIV- $b5F$ , and HIV- $b3F$  displayed no significant decrease in  $r$  or  $\Delta r$  value, under either real-time (Fig. 2) or fixed-time conditions (data not shown) ( $3'$ -processing activity was normally detected by a gel-electrophoresis method for HIV- $b5F$  and HIV- $b3F$ ; HIV- $\alpha 5F$  was not tested). The result obtained with these three ODNs indicates that anisotropy is not sensitive enough to differentiate between the DNA substrate (ds ODN 21/21) and the first reaction product (ds ODN 19/21). Furthermore, under real-time conditions, this result suggests that the processed DNA product remains tightly bound to the enzyme after  $3'$ -processing. This tight binding together with the slow catalytic step could be responsible for the observed single-turnover property of IN, even under conditions of excess DNA substrate (25). In contrast, the terminal dinucleotide is released normally from IN-DNA complexes after  $3'$ -processing, and our results indicate that the decrease in  $r$  is related to the formation of the GT product. Thus, anisotropy is reliable for monitoring the  $3'$ -processing reaction, but only if fluorescein is attached to the GT dinucleotide.

The  $3'$ -processing reaction is strongly sequence-dependent, and the endonucleolytic site includes the crucial conserved CA dinucleotide immediately preceding the GT dinucleotide. We assessed sequence specificity, using the anisotropy-based assay and the following ODNs: HIV- $\alpha 3F$  corresponds to the wild-type (wt) sequence, whereas HIV-GTGT- $\alpha 3F$  is a variant in which the  $3'$ -terminal CAGT sequence is replaced by GTGT. We also tested three nonspecific sequences, NS-TTCC- $\alpha 3F$ , NS-CACC- $\alpha 3F$ , and NS-CAGT- $\alpha 3F$ . We found that the DNA-binding step was not influenced by sequence. Indeed, the final  $r$  values obtained after DNA binding were similar for all ODNs, irrespective of sequence context, confirming that *in vitro* assays primarily reveal the nonspecific DNA binding mode of IN. In contrast, sequence had a major effect on  $\Delta r$  value in catalysis conditions (Fig. 3A). Only HIV- $\alpha 3F$  gave a large decrease in  $\Delta r$ . This decrease was abolished by replacement of the  $3'$ -terminal sequence CAGT by GTGT or the use of nonspecific sequences, even with the CAGT sequence at the  $3'$ -end. These results confirm that the CA dinucleotide is strictly required but not sufficient for activity, particularly in the presence of  $Mg^{2+}$ , which gives more stringent conditions than  $Mn^{2+}$  (4, 5). We also investigated the effect of a styrylquinoline compound, FZ41, which inhibits  $3'$ -processing activity (30). The anisotropy-based approach has been shown to be suitable for assays of the effects of inhibitors on the formation and stability of IN-DNA complexes, independently of catalysis, and has been successfully used in studies of competitive inhibitors preventing IN-DNA interactions (12). Here, we investigated both steps, DNA binding and catalysis, in the presence of inhibitor. The  $IC_{50}$  value for  $3'$ -processing was found to be  $0.74 \mu M$  (Fig. 3B), consistent with the value determined by standard gel-electrophoresis (30). In the same experiment, the study of the DNA-binding step revealed the competitive nature of FZ41:  $r_{t=0}$  decreased with increasing FZ41 concentration (not shown), suggesting that FZ41 primarily prevents the IN-DNA interaction.

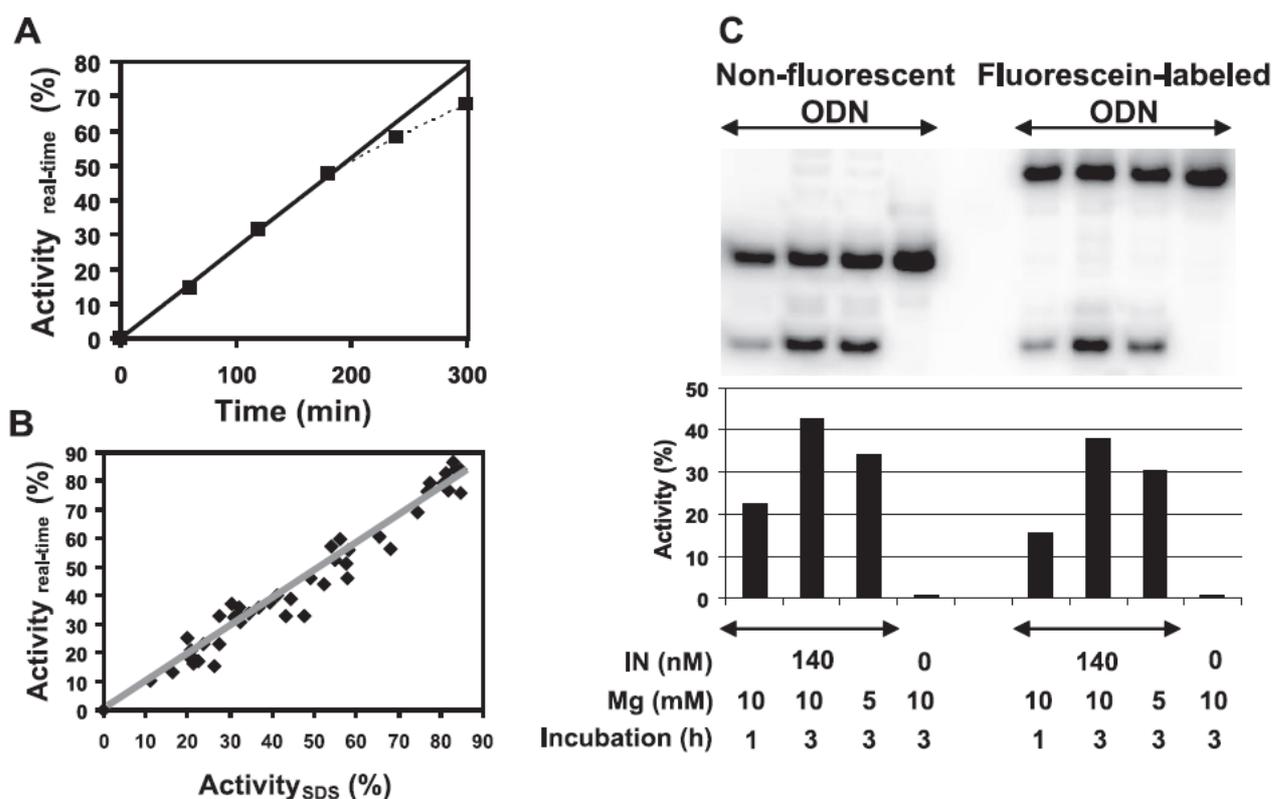
### The IN $3'$ -Processing Reaction Mechanism



**FIGURE 3. Assessment of the specificity of the  $3'$ -processing reaction as monitored by the anisotropy-based method.** A, sequence-specific decrease in steady-state anisotropy. The reaction was performed in the presence of  $Mg^{2+}$  with various DNA substrates.  $\Delta r$  (%) was plotted against time (see the legend for Fig. 2). All the ODNs tested were fluorescein-labeled on the  $3'$ -extremity of strand *a*. HIV- $\alpha 3F$  (black circles) corresponds to the wt sequence, whereas HIV-GTGT- $\alpha 3F$  (white triangles) is a variant in which the CAGT  $3'$ -terminal sequence was replaced by GTGT. Three nonspecific sequences (NS) were also studied (see "Experimental Procedures") and differed in the last four bases at the  $3'$ -end, which are explicitly mentioned in their names: NS-TTCC- $\alpha 3F$  (black squares), NS-CACC- $\alpha 3F$  (white circles), NS-CAGT- $\alpha 3F$  (white squares). The experimental conditions were similar to those described in the legend for Fig. 2. B, inhibition of  $3'$ -processing activity by FZ41 compound, as monitored by fluorescence anisotropy. Activity calculations were based on measurements of  $\Delta r_{SDS}$ .

Altogether, our data show that the decrease in  $r$  during incubation at  $37^\circ C$  is related to specific  $3'$ -processing activity and that the anisotropy-based method is reliable for inhibitor characterization. In our experiments, strand transfer products, as quantified by gel-electrophoresis, represent 5–8% of the total products, regardless of fluorescein labeling (data not shown). Only, the strand transfer of one processed DNA into a fluorescently labeled DNA (not yet processed) may underestimate the  $3'$ -processing activity determined by anisotropy. Consequently, the half-transfer reaction is more problematic at the beginning of the reaction, and its influence on  $r$  tends to disappear during  $3'$ -processing. However, this effect is probably small, particularly in the real-time assay, in which the high  $r$  values characterizing the IN-DNA complex cannot be severely influenced by strand transfer.

## The IN 3'-Processing Reaction Mechanism



**FIGURE 4. 3'-Processing activity of IN.** *A*, time course of 3'-processing activity, as obtained with the real-time assay. The DNA substrate was HIV-a3F (4 nm) in the presence of 180 nM IN and 5 mM Mg<sup>2+</sup>. Activity was calculated directly from the decrease in  $r$  values recorded at 37 °C (as shown in Fig. 2B). This activity was named "Activity<sub>real-time</sub>" in contrast to "Activity<sub>SDS</sub>" which was calculated from  $\Delta r_{SDS}$  when all the native IN-DNA complexes were disrupted by SDS after various incubation times. *B*, correlation between Activity<sub>real-time</sub> and Activity<sub>SDS</sub> values. The various activity values were obtained by varying either the incubation time or the IN:DNA ratio. We used  $r$  as measured under real-time conditions or after adding SDS to calculate the fraction of released dinucleotides, giving Activity<sub>real-time</sub> or Activity<sub>SDS</sub>, respectively. The linear fit gives a slope of 0.97 ( $R^2 = 0.96$ ). *C*, 3'-processing activity, as detected by radioactivity on a denaturing acrylamide/urea gel. *Left*, non-fluorescent DNA substrate. *Right*, fluorescein-labeled DNA substrate (HIV-a3F). The *upper* and *lower* bands correspond to the 21-mer DNA substrate and the 19-mer processed DNA product, respectively. The slower migration of HIV-a3F is due to the presence of fluorescein at the 3'-extremity.

**Quantification Analysis and Characterization of the 3'-Processing Reaction**—The 3'-processing activity of IN can be calculated by real-time assay or by quenching of the reaction with SDS. Both these approaches allow quantification of the fraction of released dinucleotides. Fig. 4A displays one example of the time course of GT release in real-time conditions. Activity increased linearly with time over ~180 min. This apparent linearity is compatible with single-turnover conditions (excess of IN over DNA substrate), when the rate constant of cleavage is low (25). Under single-turnover conditions, the time course of product formation provides a measure of the actual chemical cleavage step, which is not affected by subsequent turnovers. The kinetic profile, as determined by anisotropy measurement, gave  $k_{chemistry} = 0.0035 \text{ min}^{-1}$ , consistent with the value obtained with the standard gel-electrophoresis method ( $k_{chemistry} = 0.004 \text{ min}^{-1}$ ) (25). A similar result was obtained using the anisotropy method with SDS (data not shown). Indeed, Activity<sub>real-time</sub> and Activity<sub>SDS</sub> were well correlated (Fig. 4B). These results demonstrated that (i) all the GT dinucleotides produced by the cleavage reaction were released from the IN-DNA complexes in the real-time assay and were therefore not trapped within the complexes, (ii) the decrease in  $r$  in real-time conditions was due exclusively to the release of

dinucleotides from the IN-DNA complexes, rather than another phenomenon, such as dissociation of the enzyme from the DNA substrate or product. The single-turnover rate constant, as determined with the anisotropy-based test, was similar to that determined by gel electrophoresis, suggesting that the presence of fluorescein at the DNA 3'-extremity has no effect on DNA-binding or the subsequent catalytic step. Comparison of the 3'-processing activity of non-fluorescent and fluorescein-labeled U5-substrates by gel electrophoresis confirmed that fluorescein has no influence on activity (Fig. 4C). This result sharply contrasts with that obtained using a TAMRA-labeled ODN: this ODN was also tested by both anisotropy and gel-electrophoresis methods and, although TAMRA did not influence the number of IN-DNA complexes, 3'-processing activity was significantly decreased by a factor of 5 (data not shown). Note that the  $r$  parameter is not sensitive to the precise positioning of IN on the DNA, and this positioning may be disturbed by the TAMRA probe in the absence of a strong effect on overall DNA-binding isotherms.

**The Initial Steady-state Anisotropy Value Is Not Predictive of 3'-Processing Activity**—The final  $r$  value obtained after the DNA-binding step corresponds to the initial  $r$  value for activity ( $r_{t=0}$ ) and is related, in first approximation, to the fractional

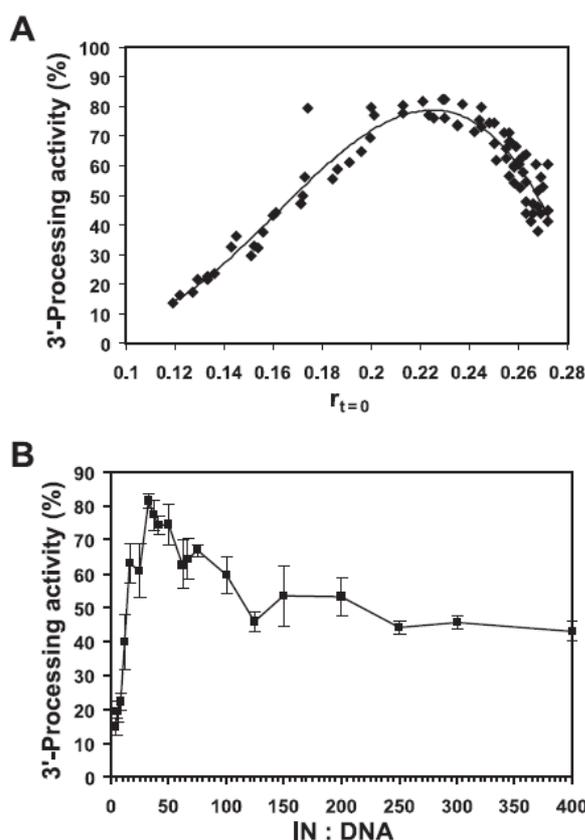


FIGURE 5. 3'-Processing activity as a function of initial steady-state anisotropy. Various mixtures of IN (from 50 to 800 nm) and HIV- $\alpha$ 3F DNA substrate (from 2 to 12 nm), giving different IN:DNA ratios, were incubated in the presence of  $Mg^{2+}$  for 20 min at 25 °C until  $r$  reached a plateau. Samples were then shifted to 37 °C and  $r_{t=0}$  (=anisotropy value before the start of the reaction) was recorded. We calculated 3'-processing activity from  $\Delta r_{SDS}$  values (at  $t = 400$  min) and plotted this activity against  $r_{t=0}$  (A) or IN:DNA ratio (B).

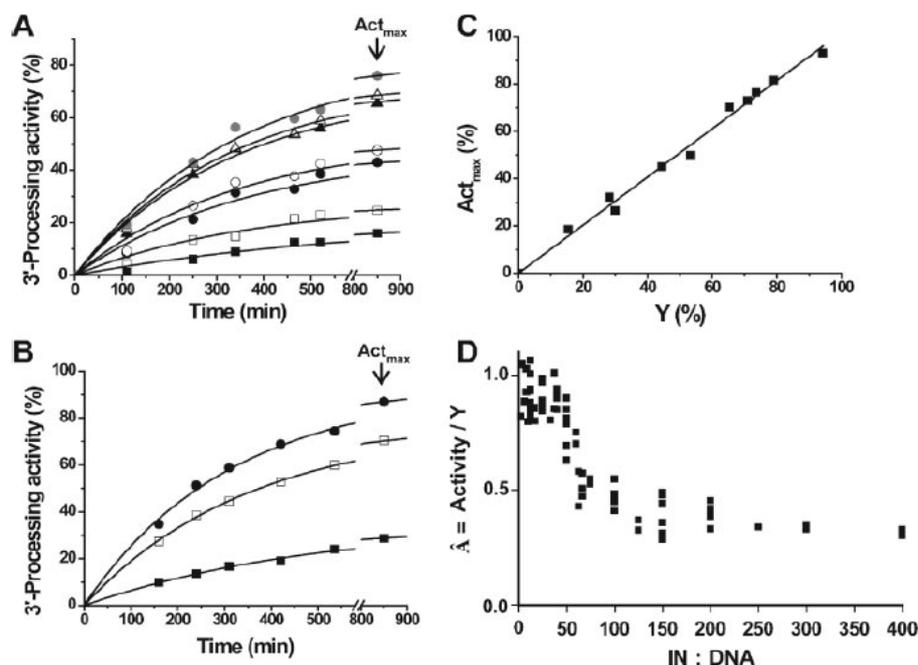
saturation function ( $Y$ ). Under single-turnover conditions, activity is expected to increase with  $Y$ , and the fraction of released dinucleotides should therefore be proportional to the number of IN·DNA complexes initially formed in the sample. Interestingly, a plot of 3'-processing activity against  $r_{t=0}$  gave a clear bell-shaped curve (Fig. 5A). The various  $r_{t=0}$  values were obtained by varying the initial IN:DNA ratio from 4.2 to 400, corresponding to different mixtures of IN (from 50 to 800 nm) and ODN (from 2 to 12 nm). The optimal ratio was found to be  $\sim 40$  (Fig. 5B) and corresponds to an  $r_{t=0}$  value of 0.226. Larger values (up to 0.272) led to a significantly decrease in the 3'-processing activity. Increases in  $r_{t=0}$  may occur for two reasons: (i) the number of free DNA molecules decreases with increasing IN:DNA ratio and (ii) self-assembly of IN on DNA may further increase the size of the complexes. The first phase ( $r_{t=0} < 0.226$ ) most likely corresponds to an increase in the number of complexes, increasing both  $r_{t=0}$  and product formation. In the second phase ( $r_{t=0} > 0.226$ ), the activity drops off suggesting that high order multimeric forms of IN on DNA, possibly aggregates, formed under conditions of high IN:DNA ratio and characterized by high  $r$  values, are less active than lower order multimeric forms. Furthermore, the  $r_{max}$  value cor-

responding to saturation ( $[DNA]_{free} \approx 0$ ) must be below 0.272, and, thus, 0.226 is the best estimate for  $r_{max}$ .

To get deeper insight into the bell-shaped phenomenon, we investigated the relationship between the fraction of product obtained at the end of the reaction and the  $Y$  parameter, which is related to the number of DNA-substrate molecules initially bound to IN (at  $t = 0$ ). The initial number of complexes was varied, and activity at  $t_{\infty}$  (corresponding to the maximum amount of released product) was determined and plotted as a function of  $Y$  (Fig. 6, A–C). The fraction of product was found to increase linearly with the initial number of IN·DNA complexes, as expected under single-turnover conditions. Optimal conditions, corresponding to an IN:DNA ratio of  $\sim 40$  (characterized by  $r = 0.226$ ), gave a final activity of 92%. This result confirms that complete saturation of the DNA substrate can be achieved at anisotropy values much lower than 0.272, and is compatible with  $r_{max} = 0.226$ . The activity per complex, specific activity ( $\text{\AA} = \text{activity}/Y$ ), was then calculated for each IN:DNA ratio (Fig. 6D). Specific activity decreased as a function of IN:DNA ratio (Fig. 6D), suggesting that aggregation occurs as soon as IN is present in excess over DNA and that the activity of large complexes (IN $_{agg}$ ·DNA) is lower than that of the lower order multimeric form (IN $_{act}$ ·DNA). At ratios exceeding 200:1, the samples contained only IN $_{agg}$ ·DNA complexes ( $r = 0.272$ ; see also TFA experiments below). However, these complexes retained significant levels of 3'-processing activity ( $\approx 35\%$  the activity of IN $_{act}$ ·DNA complexes). The number of aggregated forms of IN increased continuously with the IN:DNA ratio, suggesting that, at the peak of the bell-shaped curve, the sample already contained some IN $_{agg}$ ·DNA complexes, together with the IN $_{act}$ ·DNA complexes. From Fig. 6, we can estimate that IN $_{agg}$ ·DNA complexes in these conditions accounted for  $\sim 11$ – $15\%$  of total complexes. We made use of the additivity law of anisotropy to calculate the theoretical  $r_{max}$  value (0.218) characterizing the fully active complex (IN $_{act}$ ·DNA), in the absence of both free DNA and aggregated complexes. Thus, the bell-shaped curve shown in Fig. 5A results from two opposing trends:  $Y$  and [IN $_{agg}$ ·DNA] both increase as a function of IN:DNA ratio and are favorable and unfavorable, respectively, for 3'-processing activity.

**Characterization of the IN $_{act}$ ·DNA and IN $_{agg}$ ·DNA Complexes Using TFA**—Under site saturation conditions,  $r$  values reflect the molecular size of the complexes in a purely indicative manner. We evaluated the molecular sizes of active IN·DNA complexes in greater detail by studying representative samples corresponding to different  $r$  values, as shown in Fig. 5A, independently by TFA. This approach allows the determination of rotational correlation times ( $\theta$ ), which are directly related to hydrodynamic volume according to Equation 8. We first characterized ss and ds ODNs of various lengths (10- to 45-mer) in the absence of IN (Table 1). All ODNs were characterized by short  $\theta$  values ( $\theta_1$  and  $\theta_2$ ), which can reasonably be assigned to rotation of the fluorescein moiety around the linker at the 5'-end of the DNA and the flexibility of the linker itself. Unlike  $\theta_1$  and  $\theta_2$ , longer  $\theta$  values ( $\theta_3$  and  $\theta_4$ ) strongly depend on length and thus account for the overall rotational motion of the ODN. For each length, the longest correlation time was 2–2.7 times longer for ds than for ss ODNs. For example, the  $\theta_4$  values for

## The IN 3'-Processing Reaction Mechanism



**FIGURE 6. Product formation is related to the initial fractional saturation function.** *A*, 3'-processing kinetics for different IN:DNA ratios. Activity was measured using different ratios (below 40:1,  $r_{t=0} < 0.22$ , to limit aggregation). HIV-*a3F* concentration was 4 nM with 10 (black squares), 20 (white squares), 30 (black circles), 40 (white circles), 50 (black triangles), 60 (white triangles), or 70 nM (gray circles) IN. MgCl<sub>2</sub> concentration was 10 mM. *B*, 3'-processing kinetics in conditions of increasing ionic strength: 30 mM (black circles), 35 mM (white squares), or 40 mM MgCl<sub>2</sub> (black squares). Concentrations of ODN and IN were 4 nM and 150 nM, respectively. Activities in *A* and *B* were calculated using  $\Delta r_{SDS}$ . *C*, 3'-processing activity at  $t_{\infty}$  ( $Act_{max}$ ) as a function of initial fractional saturation function.  $Act_{max}$  corresponds to the maximal fraction of released dinucleotides for given conditions of IN:DNA ratio or MgCl<sub>2</sub> concentration (plateau in *A* and *B*) and was estimated by fitting:  $Act = Act_{max} \times (1 - e^{-k_{obs} \times t})$ . *Y* was calculated using  $r_{max} = 0.226$ . *D*, specific activity ( $\hat{A} = activity/Y$ ) as a function of the IN:DNA ratio. Total activity was normalized using fractional saturation function (*Y*), giving 3'-processing activity per complex.

**TABLE 1**  
Correlation times ( $\theta$ ) obtained for fluorescein-labeled ODNs of various lengths at 25 °C

ODN	Single-stranded <sup>a</sup>	Double-stranded
	<i>ns</i>	
10-mer	$\theta_1 = 0.13 \pm 0.02$ (0.385) $\theta_2 = 0.37 \pm 0.06$ (0.385) $\theta_3 = 1.48 \pm 0.08$ (0.23)	$\theta_1 = 0.23 \pm 0.05$ (0.48) $\theta_2 = 1.05 \pm 0.51$ (0.31) $\theta_3 = 3.30 \pm 1.17$ (0.21)
13-mer	$\theta_1 = 0.14 \pm 0.01$ (0.50) $\theta_2 = 0.38 \pm 0.02$ (0.285) $\theta_3 = 1.64 \pm 0.05$ (0.215)	$\theta_1 = 0.20 \pm 0.02$ (0.49) $\theta_2 = 0.84 \pm 0.02$ (0.26) $\theta_3 = 3.83 \pm 0.38$ (0.25)
21-mer (HIV- <i>a5F</i> )	$\theta_1 = 0.15 \pm 0.04$ (0.28) $\theta_2 = 0.42 \pm 0.08$ (0.405) $\theta_3 = 1.48 \pm 0.25$ (0.16) $\theta_4 = 4.27 \pm 0.63$ (0.155)	$\theta_1 = 0.16 \pm 0.03$ (0.425) $\theta_2 = 0.47 \pm 0.11$ (0.26) $\theta_3 = 1.75 \pm 0.56$ (0.19) $\theta_4 = 8.58 \pm 2.31$ (0.125)
36-mer	$\theta_1 = 0.12 \pm 0.03$ (0.35) $\theta_2 = 0.39 \pm 0.06$ (0.39) $\theta_3 = 1.62 \pm 0.48$ (0.17) $\theta_4 = 6.60 \pm 1.91$ (0.09)	$\theta_1 = 0.17 \pm 0.02$ (0.46) $\theta_2 = 0.65 \pm 0.17$ (0.22) $\theta_3 = 2.70 \pm 0.56$ (0.17) $\theta_4 = 18.1 \pm 2.82$ (0.15)
45-mer	$\theta_1 = 0.16 \pm 0.02$ (0.35) $\theta_2 = 0.47 \pm 0.09$ (0.40) $\theta_3 = 1.81 \pm 0.31$ (0.15) $\theta_4 = 11.1 \pm 3.06$ (0.09)	$\theta_1 = 0.26 \pm 0.03$ (0.44) $\theta_2 = 1.00 \pm 0.33$ (0.29) $\theta_3 = 4.07 \pm 0.75$ (0.17) $\theta_4 = 26.4 \pm 4.43$ (0.10)

<sup>a</sup>  $\rho_i/\rho_0$  ratios are indicated in parentheses.

the 21-mer ODN were 4.3 and 8.6 ns for ss and ds molecules, respectively. We found that the 21-mer ds ODNs HIV-*a5F*, HIV-*b3F*, HIV-*b5F*, NS-TTCC-*a3F*, and NS-CACC-*a3F* displayed similar  $\theta$  distributions and were characterized by  $r$  values of 0.045–0.055 (data not shown). In contrast, the 21-mer ds ODNs HIV-*a3F*, HIV-*aUF*, HIV-GTGT-*a3F*, and NS-CAGT-

*a3F*, which had higher  $r$  values (0.105), gave similar correlation times to the previous series, but with different  $\rho_i/\rho_0$  ratios (data not shown). For example, HIV-*a3F* had a long  $\theta_4$  (9.6 ns), similar to that of HIV-*a5F* (8.6 ns), but the flexibility of the fluorophore was significantly lower, as demonstrated by the higher  $\rho_4/\rho_0$  ratio (= 0.5) than for HIV-*a5F* ( $\rho_4/\rho_0 = 0.125$ ) (the fractional amplitude of fast motions is lower for HIV-*a3F* than for HIV-*a5F*). This decrease in flexibility seems to be related to the presence of a 3'-terminal GT and accounts for the higher  $r$  value (similar results were obtained using TAMRA; data not shown). Finally, the 21-mer ODN was the most suitable substrate for studies of IN:DNA complexes by TFA, because longer ODNs (36 or 45 bp) had a long  $\theta$ , which could complicate the analysis of IN:DNA complexes. Moreover, shorter ODNs (10 or 13 bp) were not chosen for reasons of affinity, because we found that ODN length had a strong effect on  $K_{d,app}$  (Fig. 7).

We then applied TFA to different samples corresponding to various IN:DNA ratios. We used HIV-*a5F* to prevent analytical problems due

to the activity-dependent decrease in  $r$  when fluorescein is linked to the GT dinucleotide. The activity observed at  $t_{\infty}$  for an IN:DNA ratio of 40:1 (92%) suggested that only a small amount of free ODN was present in this sample. We therefore restricted our analysis to samples corresponding to the peak of the bell-shaped curve and to the decreasing phase, so as to avoid studying mixtures of free and IN-bound ODNs. At 37 °C, two short  $\theta$  (<2.5 ns) were resolved upon the addition of IN to ODN (Table 2). The distribution in the short  $\theta$  domain was not dependent on the IN:DNA ratio. In contrast, the characteristic long  $\theta_{25\text{ °C}}$  of 8.6 ns found for free DNA was replaced by a  $\theta_{25\text{ °C}} \geq 37.8$  ns (26.8 ns at 37 °C), depending on the IN:DNA ratio (Fig. 8 and Table 2). The sample corresponding to the peak of the bell-shaped curve (IN:DNA 40:1;  $r = 0.226$ ) was rather monodisperse, as suggested by the small difference between the maximum (26.8 ns at 37 °C) and the center of gravity (31.4 ns at 37 °C) of the peak in the  $\theta$  distribution. As discussed above, samples corresponding to the TFA distribution shown in Fig. 8A contained mostly non-aggregated active forms of IN (85–89%). These complexes are characterized by a long correlation time of ~38 ns. This value is consistent with hydrodynamic properties of a dimeric form of IN on the DNA (16, 17, 31). The shift of the center of gravity toward longer  $\theta$  values probably results from the presence of small numbers of IN<sub>agg</sub>-DNA complexes. Further increases in the IN:DNA ratio displaced the

distribution toward longer  $\theta$  values (Fig. 8, B–E). Our data confirmed that the sample corresponding to the highest  $r$  value (0.272) consisted essentially of aggregated forms of IN on DNA (Fig. 8E), as demonstrated by the broad distribution above 100 ns (17).

## DISCUSSION

We used a continuous fluorescence assay to monitor the HIV-1 IN 3'-processing reaction. We found that the processed DNA product remained tightly bound to IN, thereby limiting enzyme turnover, whereas the other reaction product (GT dinucleotide) was rapidly released from the processed DNA·IN complex. This assay made it possible to characterize standard equilibrium/kinetic parameters and to obtain insight into the relationship between the size of IN·DNA complexes and catalytic activity. We studied 3'-processing under various IN:DNA ratio conditions: optimal activity was obtained at a ratio of 40:1. Higher ratios, associated with high  $r$  values, led to a decrease in product formation. The oligomeric status of active IN was further investigated by TFA. The best condition for 3'-processing activity corresponds to a majority of complexes in the sample characterized by a long correlation time of  $\sim 38$  ns. This value is compatible with a dimeric form of IN bound to one short DNA substrate mimicking a single viral DNA end. The aggregation of IN on DNA was found to be responsible for the decrease in the 3'-processing activity.

**Kinetic Analysis of 3'-Processing Activity by the Anisotropy-based Method**—The real-time assay described here is based on the use of a DNA duplex fluorescein-labeled on the terminal-3' GT. It allows quantification of the cleavage reaction resulting in physical separation of the processed 19/21-mer DNA (first product) and the fluorescently labeled dinucleotide (second

product). IN assays based on fluorescence intensity (fluorescence resonance energy transfer or fluorescence quenching) have been described in previous studies (32–35). Unlike electrophoresis-based methods, which can be used to analyze discrete time points only, fluorescence methods present the significant advantage of allowing continuous monitoring of the reaction for kinetic studies of IN in the same sample. However, only the anisotropy-based method can be used to study both the DNA-binding step and subsequent activity in the same assay. Using various DNA sequences, we showed that IN does not demonstrate highly selective binding to the viral sequence over other sequences, but, in contrast, displayed strong selectivity for catalysis (Fig. 3). This result is not subject to possible bias due to indirect comparison, as might be the case when DNA binding and catalysis are studied by different methods. It suggests that IN first binds viral DNA in a nonspecific manner and that subsequent relaxation leads to the formation of a specific and catalytically competent complex before catalysis. This relaxation step must be very slow to account for the low single-turnover rate constant as previously reported (25).

The anisotropy-based assay is reliable, because  $r$  decreased only in the presence of  $Mg^{2+}$  with the wt DNA sequence and only when fluorescein was attached to the GT 3'-extremity and was prevented by a competitive inhibitor of IN. The additivity law of anisotropy renders the quantification of activity easy and both real-time and fixed-time studies can be carried out. The real-time assay is possible, because IN·DNA complexes with DNA substrate (21/21-mer) or product (processed 19/21-mer) are highly stable in solution (Fig. 2), and the observed decrease in  $r$  is therefore due solely to the cleavage reaction and not to dissociation of the protein from DNA. Unlike many standard methods, which quantify the 19/21-mer product under denaturing conditions, the real-time assay allows quantification of the GT product under native conditions. Our kinetic analysis gave a catalytic rate of  $0.2 \text{ h}^{-1}$ , which is similar to the values published in previous studies (25, 36, 37). The strong correlation between real-time and fixed-time data (Fig. 4B) and similarities between the time courses of GT and 19/21-mer formation demonstrate that the GT dinucleotide is rapidly released from IN after the 3'-processing reaction. This release is consistent with the low apparent affinity expected for such a small ligand (see Fig. 7,  $K_d > 100 \mu\text{M}$ ). This situation is very different to that for the other reaction product (19/21-mer), which remains tightly bound to IN after catalysis. Finally, kinetic analysis showed that the catalytic rate constant was not significantly affected by the presence of fluorescein at the end of the DNA molecule, whereas inhibition was observed with TAMRA. TFA

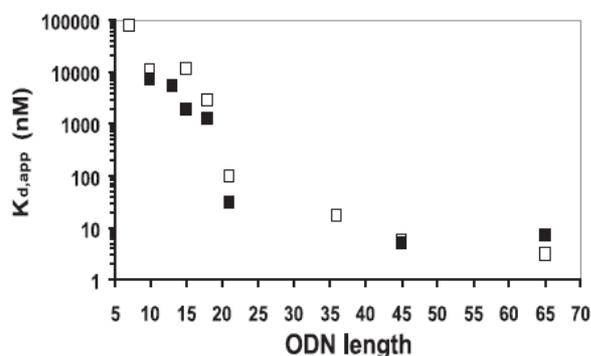


FIGURE 7. Apparent  $K_d$  values of IN as a function of DNA length (number of nucleotides). Values of  $K_{d,app}$  were determined for ss (white squares) or ds ODNs (black squares) as indicated under "Experimental Procedures."

**TABLE 2**  
Correlation times ( $\theta_{37^\circ\text{C}}$ ) obtained for IN·ODN complexes using ds ODN HIV-1a5F

Correlation times	IN:DNA <sup>a</sup>				
	40:1	67:1	100:1	200:1	400:1
$\theta_1$	0.29 (0.31)	0.19 (0.38)	0.17 (0.33)	0.15 (0.30)	0.19 (0.26)
$\theta_2$	1.93 (0.23)	2.13 (0.14)	2.14 (0.17)	2.6 (0.15)	2.5 (0.15)
$\theta_3$ , maximum	26.8 (0.46)	38.5 (0.48)	53 (0.50)	N.D (0.55)	N.D (0.59)
$\theta_3$ , center of gravity	31.4	53.3	>90	>120	>150
$\theta_{25^\circ\text{C}}^b$	37.8	54.3	74.7		

<sup>a</sup>  $\rho_i/\rho_0$  ratios are indicated in parentheses.

<sup>b</sup> Normalization of the  $\theta_3$  value for a temperature of 25 °C (according to Equation 8), allowing direct comparison with  $\theta$  values in Table 1.

## The IN 3'-Processing Reaction Mechanism

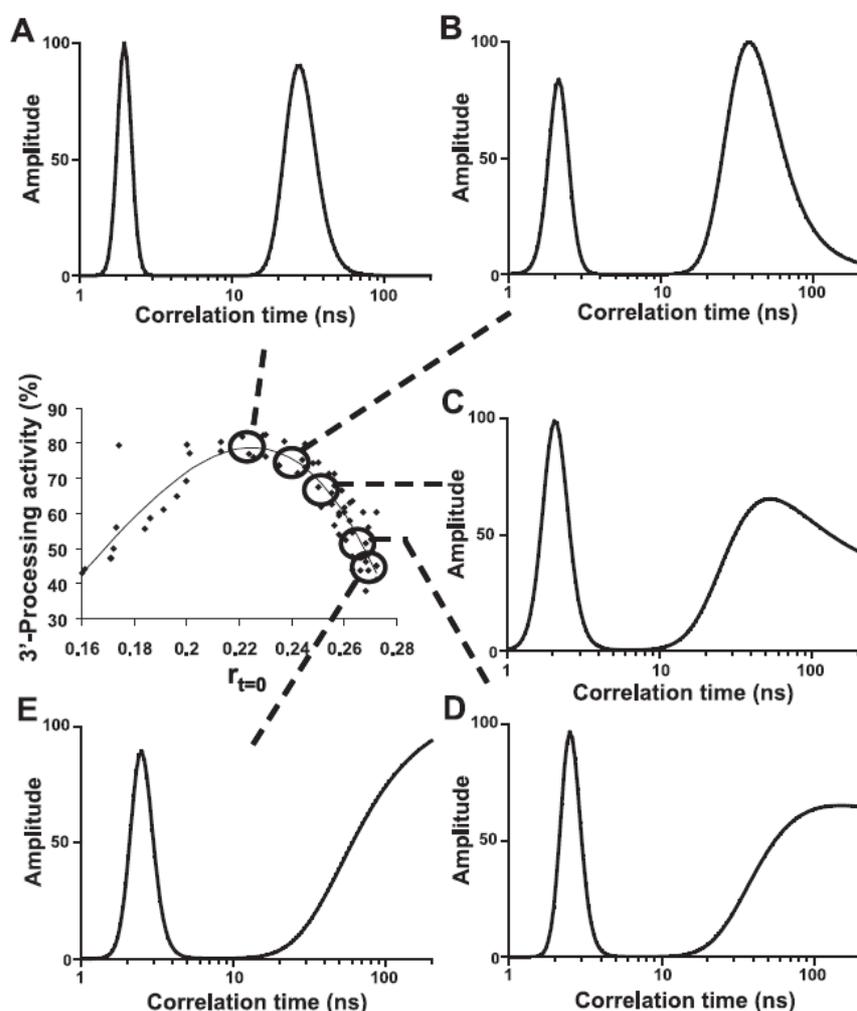


FIGURE 8. Correlation time distributions of the IN-DNA complexes obtained at different IN:DNA ratios. Representative samples corresponding to different  $r$  values in the decreasing phase of Fig. 5A were studied by TFA at 37 °C. IN:DNA ratios were 40:1 (A), 67:1 (B), 100:1 (C), 200:1 (D), and 400:1 (E). The different samples displayed similar lifetime distributions:  $\tau_1 = 0.11 \pm 0.06$  ns (26%);  $\tau_2 = 0.4 \pm 0.1$  ns (19%);  $\tau_3 = 1.9 \pm 0.1$  ns (14%);  $\tau_4 = 3.6 \pm 0.2$  ns (41%). Complete data ( $\theta$  and  $\rho_i/\rho_0$ ) are reported in Table 2.

showed that the local motion of fluorescein or TAMRA was more restricted when these fluorophores were covalently linked to a GT 3'-extremity, suggesting a direct interaction between the fluorophore and DNA. In the case of fluorescein, this interaction perturbed neither DNA binding nor cleavage. In contrast, the additional methyl groups and (or) charge effects of TAMRA may prevent the correct positioning of IN on the cleavage site or phosphodiester bond hydrolysis itself.

**Dimeric Forms of IN Are Optimal for the 3'-Processing of One DNA Extremity**—We found that the optimal conditions for 3'-processing were obtained for a majority of dimeric forms bound to the DNA ( $\theta_{\text{long}} = 38$  ns), consistent with complementation studies between individually inactive IN mutants, suggesting that at least the dimeric form is catalytically active for 3'-processing (38, 39). The optimal IN:DNA ratio (40:1) reported here corresponds to conditions used in most IN assays: IN in excess over DNA and DNA concentration maintained below the  $K_d$  value. This ratio therefore does not indicate the final stoichiometry and is highly dependent on experimen-

tal conditions. For example, increasing ionic strength shifts the  $K_d$  toward higher values and influences the optimal ratio (data not shown). Under experimental conditions in which DNA and IN concentrations were maintained above the  $K_d$  value, an optimal ratio of 2:1 was found, consistent with the stoichiometric binding of two IN protomers per DNA end (25). We and others (12, 40) have previously observed cooperative binding of IN to DNA. Consequently, using DNA concentration below the  $K_d$  value, the IN:DNA ratio determines the number of complexes formed in solution as well as the stoichiometry of these complexes. Therefore, below the optimal ratio for activity, it is very likely that, in addition to free DNA, the solution contains a mixture of monomeric and dimeric forms of IN bound to the DNA substrate. Indeed, we have previously found that, using suboptimal IN:DNA ratios (10:1), the complexes exhibit lower  $\theta$  values consistent with a monomer-dimer equilibrium at 37 °C (normalized  $\theta_{\text{long}}$ , 24–27 ns) and monomeric IN at 25 °C (normalized  $\theta_{\text{long}}$ , 16–20 ns) (16). Here, it is highly suggested that the optimal activity originates in two critical events: DNA binding and cooperativity. From this point of view, steady-state anisotropy is then particularly helpful to establish direct relationships between DNA binding parameters, the specific activity of the complexes, and their hydrodynamic properties as measured by TFA.

Our conclusion that the dimeric form of IN is active for 3'-processing under native conditions strongly parallels those obtained by Parissi and coworkers (41) with purified cross-linked multimers. The aggregation of IN on DNA was found to be strongly dependent on the IN:DNA ratio and was detrimental for 3'-processing, as previously reported for concerted integration (42). The bell-shaped curve obtained in Fig. 5 most likely originates from competition between the cooperative binding of IN on DNA leading to a dimeric form (specific IN-IN interaction, which is favored by increasing the IN:DNA ratio and enhances 3'-processing activity) and aggregation (nonspecific interactions also favored by increasing the IN:DNA ratio but detrimental for 3'-processing). The bell-shaped behavior is also compatible with the results of a recent cross-linking study (41) suggesting that monomers and tetramers display neither 3'-processing or half-strand transfer (transfer of a single viral DNA end)

activity, whereas dimers are competent for both these reactions, and only tetramers catalyze full-site concerted integration. Our results suggest that, *in vitro*, IN aggregates do not much hinder 3'-processing under IN:DNA ratio conditions leading to the formation of dimers, but that such aggregates may impede tetrameric organization in a highly competitive manner. However, TFA experiments did not indicate whether the tetrameric form on one LTR end actually existed, because the intermediate  $\theta$  distributions (Fig. 8, B–D) probably corresponded to mixtures of different multimeric forms (polydisperse solutions). It is therefore difficult to assess 3'-processing activity of tetramers, because IN:DNA ratios shifting dimer-tetramer equilibrium toward tetramers also gave rise to IN aggregation. Although recombinant IN alone is sufficient for concerted integration *in vitro* (41–44), this reaction remains much less efficient with recombinant IN than with the purified pre-integration complex (45) or virion-derived IN (46). The main limiting factor may be the multimeric organization of IN to yield the correct tetrameric form on DNA compatible with full integration process: aggregation is probably more limiting for DNA integration, which requires tetramers, than for 3'-processing, which requires dimers. Moreover, samples containing many aggregated forms displayed significant 3'-processing activity (one-third that of dimeric IN). The reason for this residual activity remains unclear.

Li and Craigie (44) recently showed that blunt-ended DNAs are more efficient substrates for concerted integration than preprocessed DNAs, whereas half-integration is more efficient with preprocessed DNA. These data, together with our results and those of Parissi *et al.* (41), suggest that dimers of IN are necessary and sufficient for 3'-processing and half-integration and that the 3'-processing step probably promotes a conformational change of IN compatible with concerted integration. This conformational change may play a key role in promoting the IN tetramerization required for concerted integration. IN is known to exist in a monomer-dimer-tetramer equilibrium in solution (16, 17, 24, 38). The multimeric form is modified on DNA, and the final quaternary structure depends on the IN:DNA ratio and probably also on the nature of the DNA substrate. From this point of view, the tetrameric organization of IN (a dimer of dimers) may be stabilized by the simultaneous presence of two LTR ends (47) and it highly suggested that the condensation-distortion properties of DNA play a key role in bringing the two ends of the viral DNA close enough in space for a dimer of dimers to form. A dimer at each LTR end, as described here, is consistent with current models based on cross-linking (48) or molecular dynamics data (49), and the 5-bp separation at target level is compatible with a tetrameric complex rather than a dimer. Several models of the tetrameric active complex exist for concerted integration: the 5-bp separation is directly compatible with a tetrameric complex displaying 2-fold symmetry, as modeled by McCammon and coworkers (31) or a tetrameric complex assuming target DNA bending (48, 50) or a hinging motion of the dimer interface (7). The formation of a dimer of dimers is probably mediated by the C-terminal domain, as suggested by the x-ray structure of the

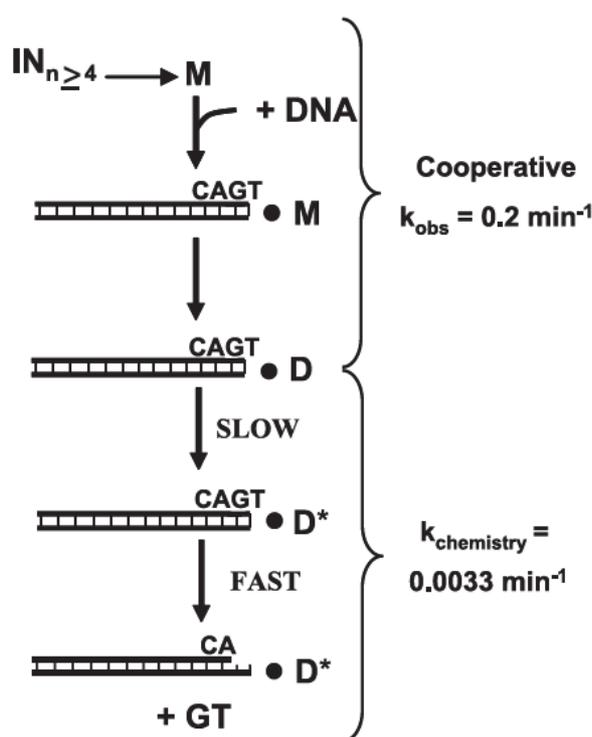


FIGURE 9. Model of the 3'-processing reaction according to binding/multimerization and catalytic parameters. Monomers (M) preferentially bind to DNA in a cooperative manner and form a dimer (D) on the LTR at one end of the DNA (see text). This model involves the conversion of large multimers ( $IN_{n \geq 4}$ ) into smaller species prior to or simultaneously with DNA binding. Once the  $IN_2$ -DNA complex has been formed, a slow  $D \rightleftharpoons D^*$  relaxation step, generating a catalytically competent complex, accounts for the very low single-turnover rate constant ( $k_{chemistry}$ ).

catalytic core/C-terminal double domain (9). Moreover, residue Trp-235 in the C-terminal domain has a significant effect on concerted integration but not on half-integration (44). The resulting symmetric organization, dispensable for 3'-processing but not for integration, may be favored or stabilized by the symmetric base sequence of the target DNA. Such symmetry in integration sites has recently been reported (51, 52).

It is noteworthy that the single-turnover rate constant of IN is particularly slow, for example about one thousandth that for restriction enzyme HincII (53). This indicates that a pre-catalysis step has a limiting effect independent from that for product release. We found that dimers were the major oligomeric forms of IN on one LTR end. IN free in solution was mostly tetrameric in the 100–200 nM concentration range and aggregated at concentrations above 250 nM (24). This suggests that the dissociation of high order multimeric forms precedes or occurs simultaneously to DNA binding, as previously reported (16) (Fig. 9). Parissi and coworkers (41) carried out a kinetic analysis of the IN-DNA complexes formation and found that the presence of one LTR induces the binding of one IN monomer (M) prior to formation of the IN dimer (D). The conversion of high order multimeric forms of IN into monomers may account for the rather slow DNA-binding step (12, 25, 27). We have previously shown that IN binds cooperatively to DNA with a Hill number compatible

## The IN 3'-Processing Reaction Mechanism

with the binding of monomers and the presence of a dimeric form of IN on DNA (12). We suggest that relaxation of the conformation of the IN·DNA complex leading to the formation of a competent and specifically bound dimeric form (D\*) is probably the limiting step accounting for the slow single turnover of 3'-processing. High IN concentrations lead to aggregation on DNA, as shown here by TFA and in other studies by means of a fluorescence correlation spectroscopy approach (54). However, we found that maximum 3'-processing activity was not controlled by absolute IN concentration but instead depended on the IN:DNA ratio. The critical IN concentration for solubility is valuable only for a given DNA substrate concentration, because aggregation properties on DNA are controlled mainly by the IN:DNA ratio. Mostly dimeric forms bound to DNA were found with, for example, a 500 nM IN solution in the presence of 12 nM DNA substrate, although DNA-free IN essentially formed aggregates at concentrations >250 nM. This does not imply that this solution is monodisperse, because IN concentration exceeds DNA concentration, but this result is consistent with the dissociation effect of DNA on high multimeric forms of IN reported in conditions of DNA excess (16). Thus, the solubility of IN may be increased by increasing DNA concentration or, alternatively, DNA binding may promote a specific conformation of IN compatible with catalysis.

*Acknowledgments*—We thank Dr. H. Leh for helpful discussions and G. Mbemba for technical assistance.

### REFERENCES

- Lutzke, R. A., Vink, C., and Plasterk, R. H. (1994) *Nucleic Acids Res.* **22**, 4125–4131
- Engelman, A., Hickman, A. B., and Craigie, R. (1994) *J. Virol.* **68**, 5911–5917
- Jenkins, T. M., Esposito, D., Engelman, A., and Craigie, R. (1997) *EMBO J.* **16**, 6849–6859
- Esposito, D., and Craigie, R. (1998) *EMBO J.* **17**, 5832–5843
- Agapkina, J., Smolov, M., Barbe, S., Zubin, E., Zatssep, T., Deprez, E., Le Bret, M., Mouscadet, J. F., and Gottikh, M. V. (2006) *J. Biol. Chem.* **281**, 11530–11540
- Mazumder, A., Neamati, N., Pilon, A. A., Sunder, S., and Pommier, Y. (1996) *J. Biol. Chem.* **271**, 27330–27338
- Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001) *EMBO J.* **20**, 7333–7343
- Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clement-Mella, C., and Mikol, V. (1998) *J. Mol. Biol.* **282**, 359–368
- Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8233–8238
- Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R., and Davies, D. R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9150–9154
- Cherepanov, P., Ambrosio, A. L., Rahman, S., Ellenberger, T., and Engelman, A. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17308–17313
- Deprez, E., Barbe, S., Kolaski, M., Leh, H., Zouhiri, F., Auclair, C., Brochon, J. C., Le Bret, M., and Mouscadet, J. F. (2004) *Mol. Pharmacol.* **65**, 85–98
- Yang, J., Xi, J., Zhuang, Z., and Benkovic, S. J. (2005) *J. Biol. Chem.* **280**, 25416–25423
- Weinberg, R. L., Veprintsev, D. B., and Fersht, A. R. (2004) *J. Mol. Biol.* **341**, 1145–1159
- Grillo, A. O., Brown, M. P., and Royer, C. A. (1999) *J. Mol. Biol.* **287**, 539–554
- Deprez, E., Tauc, P., Leh, H., Mouscadet, J. F., Auclair, C., Hawkins, M. E., and Brochon, J. C. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10090–10095
- Deprez, E., Tauc, P., Leh, H., Mouscadet, J. F., Auclair, C., and Brochon, J. C. (2000) *Biochemistry* **39**, 9275–9284
- Ciubotaru, M., Ptaszek, L. M., Baker, G. A., Baker, S. N., Bright, F. V., and Schatz, D. G. (2003) *J. Biol. Chem.* **278**, 5584–5596
- Perez-Howard, G. M., Weil, P. A., and Beechem, J. M. (1995) *Biochemistry* **34**, 8005–8017
- Bloom, L. B., Turner, J., Kelman, Z., Beechem, J. M., O'Donnell, M., and Goodman, M. F. (1996) *J. Biol. Chem.* **271**, 30699–30708
- Kuznetsova, L. G., Romanova, E. A., Volkov, E. M., Tashlitskii, V. N., Oretskaia, T. S., Krynetskaia, N. F., and Shabarova, Z. A. (1993) *Bioorg. Khim.* **19**, 455–466
- Antsyovich, S. I., Volkov, E. M., Oretskaia, T. S., Romanova, E. A., Tashlitskii, V. N., Blumenfeld, M., and Shabarova, Z. A. (1995) *Bioorg. Khim.* **21**, 774–780
- Agrawal, S., Christodoulou, C., and Gait, M. J. (1986) *Nucleic Acids Res.* **14**, 6227–6245
- Leh, H., Brodin, P., Bischerour, J., Deprez, E., Tauc, P., Brochon, J. C., LeCam, E., Coulaud, D., Auclair, C., and Mouscadet, J. F. (2000) *Biochemistry* **39**, 9285–9294
- Smolov, M., Gottikh, M., Tashlitskii, V., Korolev, S., Demidyuk, I., Brochon, J. C., Mouscadet, J. F., and Deprez, E. (2006) *FEBS J.* **273**, 1137–1151
- Brochon, J. C. (1994) *Methods Enzymol.* **240**, 262–311
- Pinskaya, M., Romanova, E., Volkov, E., Deprez, E., Leh, H., Brochon, J. C., Mouscadet, J. F., and Gottikh, M. (2004) *Biochemistry* **43**, 8735–8743
- Hazuda, D. J., Wolfe, A. L., Hastings, J. C., Robbins, H. L., Graham, P. L., LaFemina, R. L., and Emini, E. A. (1994) *J. Biol. Chem.* **269**, 3999–4004
- Vink, C., Lutzke, R. A., and Plasterk, R. H. (1994) *Nucleic Acids Res.* **22**, 4103–4110
- Zouhiri, F., Mouscadet, J. F., Mekouar, K., Desmaele, D., Savoure, D., Leh, H., Subra, F., Le Bret, M., Auclair, C., and D'Angelo, J. (2000) *J. Med. Chem.* **43**, 1533–1540
- Podtelezhnikov, A. A., Gao, K., Bushman, F. D., and McCammon, J. A. (2003) *Biopolymers* **68**, 110–120
- Lee, S. P., Censullo, M. L., Kim, H. G., Knutson, J. R., and Han, M. K. (1995) *Anal. Biochem.* **227**, 295–301
- Hawkins, M. E., Pfeleiderer, W., Mazumder, A., Pommier, Y. G., and Balis, F. M. (1995) *Nucleic Acids Res.* **23**, 2872–2880
- Wang, Y., Klock, H., Yin, H., Wolff, K., Bieza, K., Niswonger, K., Matzen, J., Gunderson, D., Hale, J., Lesley, S., Kuhlen, K., Caldwell, J., and Brinker, A. (2005) *J. Biomol. Screen.* **10**, 456–462
- Bera, S., Vora, A. C., Chiu, R., Heyduk, T., and Grandgenett, D. P. (2005) *Biochemistry* **44**, 15106–15114
- Tramontano, E., Colla, P. L., and Cheng, Y. C. (1998) *Biochemistry* **37**, 7237–7243
- Lee, S. P., Kim, H. G., Censullo, M. L., and Han, M. K. (1995) *Biochemistry* **34**, 10205–10214
- Engelman, A., Bushman, F. D., and Craigie, R. (1993) *EMBO J.* **12**, 3269–3275
- van den Ent, F. M., Vos, A., and Plasterk, R. H. (1999) *J. Virol.* **73**, 3176–3183
- Pemberton, I. K., Buckle, M., and Buc, H. (1996) *J. Biol. Chem.* **271**, 1498–1506
- Faure, A., Calmels, C., Desjobert, C., Castroviejo, M., Caumont-Sarcos, A., Tarrago-Litvak, L., Litvak, S., and Parisi, V. (2005) *Nucleic Acids Res.* **33**, 977–986
- Sinha, S., Pursley, M. H., and Grandgenett, D. P. (2002) *J. Virol.* **76**, 3105–3113
- Sinha, S., and Grandgenett, D. P. (2005) *J. Virol.* **79**, 8208–8216
- Li, M., and Craigie, R. (2005) *J. Biol. Chem.* **280**, 29334–29339
- Farnet, C. M., and Bushman, F. D. (1997) *Cell* **88**, 483–492
- Goodarzi, G., Im, G. J., Brackmann, K., and Grandgenett, D. (1995) *J. Virol.* **69**, 6090–6097
- Li, M., Mizuuchi, M., Burke, T. R., and Craigie, R. (2006) *EMBO J.* **25**, 1295–1304

## The IN 3'-Processing Reaction Mechanism

48. Gao, K., Butler, S. L., and Bushman, F. (2001) *EMBO J.* **20**, 3565–3576
49. Chen, A., Weber, I. T., Harrison, R. W., and Leis, J. (2006) *J. Biol. Chem.* **281**, 4173–4182
50. Wielens, J., Crosby, I. T., and Chalmers, D. K. (2005) *J. Comput. Aided Mol. Des.* **19**, 301–317
51. Wu, X., Li, Y., Crise, B., Burgess, S. M., and Munroe, D. J. (2005) *J. Virol.* **79**, 5211–5214
52. Holman, A. G., and Coffin, J. M. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6103–6107
53. Etzkorn, C., and Horton, N. C. (2004) *Biochemistry* **43**, 13256–13270
54. Vercammen, J., Maertens, G., Gerard, M., De Clercq, E., Debysier, Z., and Engelborghs, Y. (2002) *J. Biol. Chem.* **277**, 38045–38052



## I.C. Conclusion

Nous avons mis au point un test d'activité basé sur l'anisotropie statique de fluorescence qui permet de suivre en temps réel la réaction de 3'-processing catalysée par l'IN du VIH-1. Nous avons pu observer tout d'abord que l'ADN processé, le premier produit de la réaction, reste lié à l'IN avec une affinité comparable à celle de l'ADN substrat. Ces deux affinités comparables peuvent représenter un facteur limitant à la cinétique de turn-over de l'IN du VIH-1. Le dinucléotide GT, le deuxième produit de la réaction, est quand à lui libéré dans le milieu réactionnel au fur et à mesure de sa production. Ce test présente deux avantages lorsqu'on le compare aux autres tests d'activité de 3'-processing de l'IN du VIH-1. Tout d'abord, il permet d'observer la fixation de l'IN du VIH-1 sur l'ADN viral et l'activité de coupure résultante sur un même échantillon. Il permet en plus de caractériser les paramètres d'équilibre et cinétique relatifs à chacune de ces deux étapes.

Ce test d'activité nous a permis de mettre en évidence deux phases selon le ratio IN / ADN utilisé. Pour les ratio faibles ( $\leq 40$ ), les valeurs d'anisotropies mesurées après la fixation, directement reliées à la fraction de saturation de l'IN, sont prédictives de l'activité finale de coupure endonucléolytique de l'ADN viral. Ces résultats confirment les résultats obtenus par Smolov et al. (Smolov *et al.*, 2006) qui montrent que l'IN catalyse la réaction de 3'-processing selon un mécanisme de « single turn-over ». Ce mécanisme peut s'expliquer par le caractère intermédiaire de la réaction de 3'-processing dans le processus global d'intégration de l'ADN viral dans l'ADN cellulaire dont le produit de cette première réaction est le substrat de la seconde, l'intégration concertée. Par conséquent, les affinités de l'IN pour l'ADN processé sont équivalentes à celles de son substrat. Pour les ratio plus élevés ( $\geq 40$ ), l'anisotropie des complexes IN / ADN continue d'augmenter, mais l'activité résultante montre une diminution significative de 85% à 35% d'ADN processé. Ces différents complexes ont été caractérisés par anisotropie de fluorescence résolue en temps. Les temps de corrélation calculés nous ont permis de montrer que pour un ratio de 40, qui correspond aux conditions d'activité maximale des complexes IN / ADN, l'IN est majoritairement fixée à l'ADN sous forme dimérique. Lorsque le ratio augmente, les temps de corrélation calculés passent de manière transitoire de 38 ns (temps de corrélation calculé pour un dimère d'IN du VIH-1 fixé sur l'ADN) à des valeurs supérieures à 90 ns (caractéristiques de la présence d'agrégats). Ce phénomène est symptomatique d'une IN qui agrège sur l'ADN, avec pour les

temps de corrélation intermédiaires un mélange comprenant des complexes dimériques et des agrégats dont la proportion augmente avec le ratio IN / ADN.

Nos résultats en regard avec ceux obtenus par d'autres équipes ((Li *et al.*, 2005); (Faure *et al.*, 2005)), nous ont permis de construire un modèle mettant en évidence les mécanismes mis en jeu durant le 3'-processing en ce basant sur les paramètres de fixation / multimérisation en présence d'ADN et catalytiques de l'IN du VIH-1. L'IN libre dans le milieu réactionnel est présente sous forme de complexes tétramériques lorsque sa concentration est de l'ordre de 100 à 200nM, au-delà de ces concentrations, l'état multimérique est d'ordre supérieur à quatre ou agrégé (Leh *et al.*, 2000). La fixation à l'ADN présente une cinétique lente de fixation avec un effet de coopérativité (Deprez *et al.*, 2004) qui peut s'expliquer par une fixation séquentielle de l'IN. Dans un premier temps, les multimères d'IN en solution se dissocient en monomères, puis un monomère d'IN se fixe sur l'ADN, un deuxième monomère pourra alors se fixer de manière facilitée sur le complexe IN monomérique / ADN. Néanmoins, nous ignorons s'il s'agit d'une dissociation des formes à haut poids moléculaires provoquée par la présence d'ADN (des expériences précédentes suggèrent ce modèle (Deprez *et al.*, 2001)) ou bien d'un déplacement d'équilibre. Dans ce dernier cas, en l'absence d'ADN, les formes monomériques seraient minoritaires. Ceci permet d'une part d'expliquer la cinétique lente de fixation de l'IN sur l'ADN (Smolov *et al.*, 2006) et d'autre part, cette hypothèse est corroborée par les résultats de cross-link résolu en temps qui font état d'une fixation séquentielle de l'IN sur l'ADN : les espèces monomériques liées à l'ADN précèdent les espèces dimériques (Faure *et al.*, 2005). Ce résultat est en accord avec l'indice de coopérativité ( $\tilde{n} = 2$ ) (Deprez *et al.*, 2004). Dans un deuxième temps, ce complexe IN dimérique / ADN est rendu catalytiquement actif par un mécanisme non identifié à ce jour qui est limitant pour l'ensemble de la cinétique apparente de la coupure endonucléolytique. Ce mécanisme lent pourrait être un réarrangement soit structural soit spatial du dimère d'IN sur l'ADN viral mais également un changement conformationnel dans le site actif du monomère d'IN responsable de la coupure endonucléolytique.

## ***II. Etude du mécanisme de reconnaissance IN / ADN : un mode d'interaction à l'ADN spécifique mis en évidence par une IN marquée avec une technique enzymatique***

### **II.A. Introduction**

L'IN du VIH-1 possède des propriétés agrégatives telles que ses études biochimiques et structurales en sont sévèrement ralenties. A ce jour, il n'existe aucune donnée structurale concernant l'IN dans son intégralité seule ou en présence d'ADN. Les seules données que nous possédons sont les structures de chacun des trois domaines ainsi que des deux doubles domaines IN<sup>1-212</sup> et IN<sup>55-288</sup>, et pour chacun d'eux sans ADN. De ce fait, les spécificités de reconnaissance de l'interaction IN / ADN viral lors du 3'-processing restent peu comprises. La solubilité de l'IN du VIH-1 a certes été améliorée que se soit par des mutations ponctuelles ou par l'ajout de détergents, mais ces deux modifications ne sont pas forcément sans effet sur l'activité de 3'-processing en présence de Mg<sup>2+</sup>. En effet, certains détergents inhibent l'activité de 3'-processing Mg<sup>2+</sup> dépendante et un mutant de solubilité présente une résistance aux diketo acides. De plus, la mutation F185K, qui est la mutation de solubilité la plus utilisée, semble abolir complètement l'intégration concertée du génome du VIH-1 dans le cadre d'infection de cellules en cultures. Ces différents résultats expérimentaux montrent que la présence de ces mutations ou de détergent dans le milieu réactionnel peuvent modifier les interactions IN / IN lors des phénomènes de multimérisation, mais également toucher les interactions IN / ADN dans le cadre des différentes activités catalytiques de l'IN. Ceci implique que leur présence dans les structures cristallographiques pourrait introduire des artefacts non négligeables sur l'analyse de ces données et sur les conclusions résultantes.

Le clivage endonucléolytique réalisé lors du 3'-processing nécessite la présence d'une séquence consensus CA précédant les deux nucléotides libérés. D'autres bases semblent primordiales pour l'activité dans un contexte Mg<sup>2+</sup> dépendant, cependant *in vitro*, aucune reconnaissance spécifique des séquences U5 ou U3 processées n'a été mise en évidence pour l'IN du VIH-1. A ce jour il est admis que la spécificité de l'IN ne s'exprime que lors des événements catalytiques. Cependant, de par sa faible solubilité, il est difficile de séparer les propriétés propres aux intégrases rétrovirales de celles dues à la propension à agréger. Nous avons eu, au cours de nos recherches, l'opportunité de travailler sur une autre IN rétrovirale : l'IN de PFV-1 (Primate foamy virus de type 1). Cette dernière présente un profil de solubilité plus clément que celui de l'IN du VIH-1. Il existe un certains nombre de différences non

négligeables lorsque l'on considère le cycle de réplication de PFV-1, qui est un spumavirus, comparé au cycle des lentivirus : présence d'une transcription inverse tardive, un 3'-processing asymétrique qui n'a lieu qu'au niveau de l'extrémité U5, des interactions avec les partenaires cellulaires différentes notamment avec LEDGF/p75 qui n'interagit qu'avec les intégrases lentivirales. Mais l'étape d'intégration est également une étape obligatoire du cycle de réplication des spumavirus et l'IN de PFV-1 présente des homologies structurales (une organisation en trois domaines) et biochimiques (quatre activités reproduites *in vitro* : 3'-processing, transfert de brin, désintégration et coupure endonucléolytique de la séquence palindromique correspondant à la jonction des cercles 2-LTR) avec l'IN du VIH-1.

Nous avons utilisé différentes techniques physico-chimiques (Spectroscopie de Corrélation à Fluorescence (FCS), chromatographie d'exclusion, Anisotropie résolue en temps (TFA)) afin de caractériser l'état multimérique de l'IN de PFV-1 seule en solution et en présence d'ADN viral mimant l'extrémité 3'U5 du LTR de l'ADN viral. Les paramètres d'équilibre et catalytique relatifs à la fixation sur l'ADN viral et au 3'-processing de l'IN de PFV-1 ont été déterminés avec le test d'activité de 3'-processing basé sur l'anisotropie de fluorescence statique (Guiot *et al.*, 2006). Son profil de solubilité plus clément que celui de l'IN de HIV-1 a permis de réaliser son marquage par un petit fluorophore (le TAMRA) au niveau d'une séquence spécifique ajoutée au niveau de son extrémité C-terminale par l'action d'une transglutaminase. Outre les expériences de FCS, cette IN marquée nous a servi à mener des expériences de FRET entre l'ADN marqué à la fluorescéine et le TAMRA fixé sur l'IN de manière à étudier les spécificité d'interactions IN / ADN. Toute cette étude des paramètres catalytiques et structuraux de l'IN de PFV-1 nous a permis de différencier les propriétés intrinsèques des intégrases rétrovirales de celles dues à leur propension à agréger.

## **II.B. Manuscrit**

Ce manuscrit a été publié dans « the Journal of Biological Chemistry », vol 283 (41), pp : 27838-49, le 10 octobre 2008.

# Insight into the Integrase-DNA Recognition Mechanism

## A SPECIFIC DNA-BINDING MODE REVEALED BY AN ENZYMATICALLY LABELED INTEGRASE\*

Received for publication, April 29, 2008, and in revised form, July 10, 2008. Published, JBC Papers in Press, August 12, 2008, DOI 10.1074/jbc.M803257200

Olivier Delelis<sup>1</sup>, Kevin Carayon<sup>1</sup>, Elvire Guiot, Hervé Leh, Patrick Tauc, Jean-Claude Brochon, Jean-François Mouscadet, and Eric Deprez<sup>2</sup>

From the Laboratoire de Biologie et Pharmacologie Appliquée, CNRS, Ecole Normale Supérieure Cachan, Institut d'Alembert, 61 Ave. du Président Wilson, 94235 Cachan, France

Integration catalyzed by integrase (IN) is a key process in the retrovirus life cycle. Many biochemical or structural human immunodeficiency virus, type 1 (HIV-1) IN studies have been severely impeded by its propensity to aggregate. We characterized a retroviral IN (primate foamy virus (PFV-1)) that displays a solubility profile different from that of HIV-1 IN. Using various techniques, including fluorescence correlation spectroscopy, time-resolved fluorescence anisotropy, and size exclusion chromatography, we identified a monomer-dimer equilibrium for the protein alone, with a half-transition concentration of 20–30  $\mu\text{M}$ . We performed specific enzymatic labeling of PFV-1 IN and measured the fluorescence resonance energy transfer between carboxytetramethylrhodamine-labeled IN and fluorescein-labeled DNA substrates. FRET and fluorescence anisotropy highlight the preferential binding of PFV-1 IN to the 3'-end processing site. Sequence-specific DNA binding was not observed with HIV-1 IN, suggesting that the intrinsic ability of retroviral INs to bind preferentially to the processing site is highly underestimated in the presence of aggregates. IN is in a dimeric state for 3'-processing on short DNA substrates, whereas IN polymerization, mediated by nonspecific contacts at internal DNA positions, occurs on longer DNAs. Additionally, aggregation, mediated by nonspecific IN-IN interactions, occurs preferentially with short DNAs at high IN/DNA ratios. The presence of either higher order complex is detrimental for specific activity. Ionic strength favors catalytically competent over higher order complexes by selectively disrupting nonspecific IN-IN interactions. This counteracting effect was not observed with polymerization. The synergic effect on the selection of specific/competent complexes, obtained by using short DNA substrates under high salt conditions, may have important implications for further structural studies in IN-DNA complexes.

Retroviral integrases (INs)<sup>3</sup> catalyze the integration of viral DNA into the host genome ensuring its perpetuation in the host

cell. The integration process requires two catalytic steps. During the first step, titled 3'-processing (3'-P), IN specifically removes two nucleotides from each viral DNA end. IN then transfers both ends into target DNA by a one-step transesterification reaction, resulting in full-site integration. IN alone is competent for the insertion process (1). Retroviral INs consist of three functional domains. The core domain contains the DDE catalytic triad and is flanked by the N-terminal domain (involved in multimerization) and the C-terminal DNA-binding domain. The integrity of the DDE triad and a metallic cofactor are strictly required for enzymatic activity.

Biochemical and structural studies of HIV-1 IN have been severely impeded because of solubility. No structural data are available for the full-length protein (free or bound to DNA) to date, although x-ray structures of DNA-free single or double domains have been successfully solved (2–5). Hence, the sequence specificity of the 3'-P reaction remains poorly understood and little is known about the mechanism of IN/DNA substrate recognition. Point mutations or the use of detergent may improve the solubility but only to a limited extent (6, 7) and may also cause changes to IN properties. For instance, detergent is detrimental to 3'-P activity in the presence of the physiological cofactor  $\text{Mg}^{2+}$  (8), and one soluble mutant was found to be resistant to diketo-acid compounds (9).

Endonucleolytic cleavage depends primarily on the presence of the canonical CA sequence preceding the removed dinucleotide. Other positions are crucial for 3'-P activity under  $\text{Mg}^{2+}$  conditions, despite the absence, *in vitro*, of IN preference/specificity for the cognate U5 or U3 terminal sequence at the DNA binding level (10–12). Thus, it is generally hypothesized that IN specificity is fully explained at the catalytic level. However, due to poor solubility, it is a difficult task to determine which enzymatic properties of HIV-1 IN actually correspond to intrinsic properties of retroviral INs and which properties are related to aggregation. During the course of our study on HIV-1 IN, we have observed significant differences relating to solubility between HIV-1 IN and another retroviral IN, primate foamy virus-1 (PFV-1) IN. PFV-1 is the prototype of foamy viruses belonging to the retrovirus family and differ from lentiviruses in some aspects. For instance, PFV-1 reverse transcription

cence resonance energy transfer; HIV-1, human immunodeficiency virus, type 1; LTR, long terminal repeat; ODN, oligodeoxynucleotide; PFV-1, primate foamy virus type-1; TFA, time-resolved fluorescence anisotropy; TGase, transglutaminase; 3'-P, 3'-processing; Mo, monomeric; Di, dimeric; Te, tetrameric; TAMRA, carboxytetramethylrhodamine.

\* This work was supported by TrioH European Project Grant FP6 503480, Agence Nationale de la Recherche Grant 06-PCVI-0015, and grants from CNRS and Institut d'Alembert. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> Both of these authors contributed equally to this work.

<sup>2</sup> To whom correspondence should be addressed. Tel: 33-147-40-23-94; Fax: 33-147-40-76-84; E-mail: deprez@lbpa.ens-cachan.fr.

<sup>3</sup> The abbreviations used are: IN, integrase; IN<sup>T</sup>, TAMRA-labeled integrase; FCS, fluorescence correlation spectroscopy; Fl, fluorescein; FRET, fluores-

occurs at late stages of the replication cycle (13, 14). Interaction with host cellular cofactors is also distinct, since LEDGF/p75 only interacts with lentiviral INs (15). Furthermore, only the U5 end is processed due to the asymmetric nature of the integration process in the case of PFV-1 (16). However, integration is an obligatory step for replicating PFV-1, a hallmark of retrovirus life cycle (16). PFV-1 IN shares common structural features with other retroviral INs, such as three-domain organization as well as the three major catalytic activities, 3'-P, strand transfer, and disintegration (17–19).

To get deeper insight into the IN/DNA recognition mechanism, we studied and compared solubility, oligomeric status and catalytic efficiency of PFV-1 and HIV-1 IN. We used various methods, such as time-resolved fluorescence anisotropy (TFA), fluorescence correlation spectroscopy (FCS), and size exclusion chromatography, and found that free PFV-1 IN was significantly more soluble than HIV-1 IN. This enabled us to successfully apply transglutaminase (TGase)-mediated TAMRA labeling to a soluble IN at a single predefined position. Fluorescence resonance energy transfer (FRET) experiments and a DNA-binding anisotropy-based assay were conducted to assess the specificity of IN-DNA interactions. A specific mode of interaction was clearly shown for PFV-1 IN but not for HIV-1 IN. DNA promotes the catalytically competent dimeric state. The dimerization process is in competition with other processes (which were found to be detrimental to activity), such as IN polymerization and aggregation, mediated by nonspecific IN-DNA and IN-IN interactions, respectively. A selection of specific/competent complexes may be achieved using short DNA substrates under conditions of high ionic strength.

## EXPERIMENTAL PROCEDURES

**Purification and IN Labeling by Guinea Pig TGase**—For DNA-binding and 3'-P assays, unlabeled wild-type HIV-1 and PFV-1 IN were purified as previously described (8). The plasmid encoding PFV-1 IN with the specific C-terminal PKPQQFM tag for TGase-mediated labeling was constructed by PCR amplification of pET15b-PFV-IN (20) with P1 and P2 primers (P1, 5'-ACA TAT GTG TAA TAC CAA AAA ACC AAA CCT GG-3'; P2, 5'-AGG ATC CTA CAT AAA CTG CTG AGG TTT TGG CTC GAG TTC ATT TTT TTC-3'). The corresponding plasmid for tagged HIV-1 IN was constructed by amplification of pET15b-HIV-IN (8) with P3 and P4 primers (P3, 5'-CCA TAT GTT TTT AGA TGG AAT AGA TAA-3'; P4, 5'-CGG ATC CTA CAT AAA CTG CTG AGG TTT TGG GTC CTC ATC CTG TCT ACT-3'). The underlined nucleotides correspond to the TGase tag (PKPQQFM) located at the C terminus of IN. The resulting DNAs encoding IN-PKPQQFM were inserted into pET15b.

For labeling, IN attached to Ni<sup>2+</sup>-beads was incubated overnight at 4 °C with 0.5 mM TAMRA cadaverine (Molecular Probes) and 1 unit of guinea pig TGase (Sigma) in the labeling buffer (20 mM Hepes, pH 7.5, 10 mM CaCl<sub>2</sub>). Beads were extensively washed with Tris buffer (50 mM, pH 8, 1 M NaCl) to remove free fluorophores. Elution was done with 1 M imidazole. TAMRA-labeled or unlabeled PFV-1 IN was further purified by cation exchange chromatography on a Mono S column (Bio-Rad) previously equilibrated with 20 mM Hepes, pH 7.5, 0.1 M

NaCl. IN was diluted before loading (1 ml/min) and a linear salt gradient (0.1–1 M NaCl) was applied at 1 ml/min. IN was eluted at 450–550 mM NaCl (cation exchange chromatography was unsuccessful with HIV-1 IN for solubility reasons). It was possible to further concentrate PFV-1 IN using Amicon Ultra centrifugal filters (Millipore). Concentrations up to 250 μM in 20 mM Hepes, pH 7.5, 75 mM NaCl were obtained. Ionic strength was adjusted by buffer exchange during the concentration step. Dialysis was not used, since protein aggregation was reproducibly observed during the dialysis step of concentrated samples (significant precipitation occurred at PFV-1 IN concentrations of 12 mg/ml, whichever protocol was used).

**Size Exclusion Chromatography**—100 μl of varying concentrations of purified PFV-1 IN (from 2 to 60 μM) were applied at room temperature onto a Superdex 200 preparation grade column (Hiload 16/60; Amersham Biosciences), using an AKTA FPLC system (Amersham Biosciences) previously equilibrated with the protein buffer (20 mM Tris-HCl, pH 8, 0.05 or 0.5 M NaCl, 10% glycerol, 50 μM ZnSO<sub>4</sub>, and 4 mM β-mercaptoethanol). The column was calibrated with the following globular proteins: sweet potato β-amylase (200 kDa), yeast alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (29 kDa), and horse heart cytochrome *c* (12.5 kDa) (MW-GF-200 kit; Sigma). Fractions were collected at a flow rate of 0.3 ml/min, under a pressure of 0.5 megapascals. Absorbance was measured at 280 nm. The exclusion volume of the column  $V_0$  was measured by calibration with blue dextran (2000 kDa).

**Characterization of TAMRA-labeled IN (IN<sup>T</sup>)**—We evaluated the number of TAMRAs per labeled protein by mass spectrometry analysis; IN<sup>T</sup> and the unlabeled tagged IN were subjected to SDS-PAGE. After gel staining, the gel slides containing the corresponding proteins were destained in 25 mM NH<sub>4</sub>CO<sub>3</sub>, 50% acetonitrile. A proteolysis using sequence grade trypsin (Promega) was then performed overnight at 37 °C. The resulting peptides were recovered by the addition of 50% acetonitrile, 5% trifluoroacetic acid and dried in a SpeedVac. They were reconstituted in 5% acetonitrile, 0.1% trifluoroacetic acid and applied on a SEND-ID protein array according to Ciphergen's protocol. The mass spectra were obtained by reading the protein chip using a SELDI mass spectrometer. The resulting spectra were compared with the *in silico* profile obtained with GPMW software (Lighthouse Data).

The labeling yield of purified IN<sup>T</sup> was estimated by both absorbance spectroscopy and FCS. Given the molar extinction coefficient of TAMRA-cadaverine at 556 nm (62,000 M<sup>-1</sup>cm<sup>-1</sup>) and the results from the Bradford protein assay, the labeling yield was estimated to be 45% by absorbance spectroscopy. The FCS autocorrelation curve (Fig. 1E) was also used to estimate the labeling yield, since it is possible to deduce from the inverse autocorrelation amplitude the mean number of labeled molecules in the detection volume. Here,  $g(0)^{-1}$  was equal to 0.025 and the detection volume was 1.27 μm<sup>3</sup> (the lateral  $\omega_0$  and axial  $z_0$  dimensions are equal to 0.45 and 1.5 μm, respectively) giving an estimated concentration of IN<sup>T</sup> (52 nM). Considering the concentration of total IN (100 nM), the deduced labeling yield (~50%) is in good agreement with that determined by absorbance spectroscopy. This calculation is

## Heterogeneity and Catalytic Properties of IN-DNA Complexes

evidently correct only if a single fluorophore is attached to the protein (SELDI analysis of IN<sup>T</sup> confirmed that a single TAMRA moiety was covalently and specifically attached to the protein C-terminal end).

IN<sup>T</sup>, as obtained after the labeling procedure, was also tested for 3'-P at the U5 end (21-mer oligonucleotide) using a standard assay (8) and was compared with both the untagged and the unlabeled tagged proteins. 3'-P and half-transfer products were about 23–27% and 3.5–4.5%, respectively, for a 3-h incubation (data not shown), and no significant difference was observed between the three samples, showing that neither the PKPQQFM tag nor the attached fluorophore influences the catalytic activity.

**PFV-1 ODNs for DNA Binding/3'-P Anisotropy-based Assays and FRET Experiments**—All ODNs (≤65-mer) were purchased from Eurogentec (Liège, Belgium) and further purified by gel electrophoresis. ODNs of various lengths (15–65-mer) were tested for 3'-P activity, all mimicking the specific (cognate) terminal U5 sequence: PFV-CAAT (sequence of the 65-mer, 5'-GAA CTA CAC TTA TCT TAA ATG ATG TAA CTC CTT AGG ATA ATC AAT ATA CAA AAT TCC ATG ACA AT-3'. The 3'-dinucleotide cleaved by IN is underlined). The processed strand and its complementary nonprocessed strand were denoted A and B, respectively. Fluorescein (Fl) was attached at the 5'- or 3'-end of strand A for DNA-binding experiments. Fl was attached at the 3'-end of strand A for the activity assay. The PFV-GTAT sequence was identical, except that the four 3'-term bases of strand A were GTAT instead of the canonical sequence CAAT. The specific HIV-1 ODN (U5 sequence) is provided in Ref. 21. The three nonspecific (random) sequences used for competition experiments (Fig. 4) were as follows: 5'-ACA TAA TCT AAA ATA ATT GCC-3' (21-mer), 5'-ACC TAT GCG CCG CTA GAT TCC-3' (21-mer), and 5'-TCA AGC TAG AAG ATT ATC TCA AGT ACA TAA TCT AAA ATA ATT GCC-3' (45-mer). Annealing of oligonucleotides was obtained by mixing equimolar amounts of complementary strands in 20 mM Hepes, pH 7.5, 100 mM NaCl, heating to 85 °C for 5 min, and slow cooling to 25 °C.

For FRET experiments, ODNs of various lengths (21–300-mer) were studied, all mimicking the PFV-1 U5 end. The double-stranded ODN was Fl-labeled at the 5'-end (A or B strand) (Fig. 3A). Fl-labeled 100- and 300-mer double-stranded ODNs were obtained by PCR. All PCRs were performed using the pHSRV13 plasmid encoding the proviral PFV-1 genome. The 100-mer *BF15* and 300-mer *BF15* were obtained using, as primers, the single-stranded 21-mer *BF15* ODN and a second ODN hybridizing to an internal sequence located, respectively, 100 or 300 bp from the U5 extremity. The 100-mer *AF15* and 300-mer *AF15* were obtained using, as primers, the nonfluorescent single-stranded 21-mer B ODN and a 5'-Fl-labeled ODN hybridizing to an internal sequence located, respectively, 100 or 300 bp from the U5 extremity. After amplification, PCR products were purified on agarose gel using the Qiagen gel extraction kit.

**Steady-state Fluorescence Anisotropy-based Assay**—Anisotropy-based assay is a quantitative method to determine both DNA binding and catalytic parameters of IN using the same sample (21). Briefly, IN binding to Fl-labeled DNA increases

the steady-state anisotropy value ( $r$ ) (measured on a Beacon instrument (Panvera, Madison, WI)), allowing the calculation of the fraction of DNA sites bound to IN. The DNA-binding step can be recorded at 25 °C using Fl-labeled ODNs at the 5'- or 3'-end of either the A or B strand. By shifting the temperature to 37 °C, the activity-dependent decrease in the  $r$  value allows quantification of the 3'-P reaction. This occurs only if Fl is initially linked to the AT dinucleotide (the released PFV-1 dinucleotide). The fraction of released dinucleotides ( $F_{\text{dinu}}$ ) can be calculated using the relative decrease of  $r$  compared with the initial value obtained at the end of the DNA-binding step ( $r_{t=0}$ ) (real time assay) (21). No decrease in the  $r$  value was observed with the PFV-GTAT control sequence, in contrast to the PFV-CAAT canonical sequence (data not shown). The activity can be also calculated in fixed time experiments (3'-P is stopped by SDS),  $F_{\text{dinu}} = (r_{\text{NP}} - r)/(r_{\text{NP}} - r_{\text{dinu}})$ , where  $r_{\text{NP}}$  and  $r_{\text{dinu}}$  are the anisotropies for pure solutions of nonprocessed double-stranded ODN and dinucleotide, respectively ( $r_{\text{dinu}}$  was measured using the 5'-AT-3'-Fl dinucleotide).

The formation of IN-DNA complexes and the subsequent 3'-P catalytic reaction were performed by incubating Fl-labeled ODNs with unlabeled IN in buffer A (20 mM Hepes, pH 7.5, 1 mM dithiothreitol, 10 mM MgCl<sub>2</sub>). IN, DNA, and NaCl concentrations are explicitly mentioned in the figure legends.

**FRET Experiments**—FRET between Fl-labeled DNA (donor) and TAMRA-labeled IN (acceptor) was monitored with an Eclipse (Varian) spectrofluorimeter. DNA concentration was constant (20 nM) in buffer A + 50 mM NaCl. IN<sup>T</sup> was progressively added, and the decrease in the steady-state emission intensity of the donor was recorded at 25 °C. The excitation wavelength was 500 nm. The donor quenching ( $q_D$ ) was measured at 520 nm.  $q_D$  was estimated using the equation,  $q_D = 1 - (I_{D,[IN^T]}/I_{D,0})$ , where  $I_{D,0}$  and  $I_{D,[IN^T]}$  represent the donor emission intensity at 520 nm in the absence (DNA alone) and in presence of acceptor (along the titration), respectively. Excitation and emission slit widths were 10 and 5 nm, respectively. For each DNA/IN<sup>T</sup> mixture, a corresponding control experiment was done using unlabeled IN to check the signal stability.

**Fluorescence Correlation Spectroscopy**—FCS measurements were performed under two-photon excitation (840 nm) on a home-built system using an 80-MHz mode-locked Tsunami Ti:Sapphire laser (pulse 100 fs) pumped by a Millennia solid state laser (Spectra Physics) and a Nikon TE2000 inverted microscope. Before entering through the epifluorescence port of the microscope, the laser beam was expanded with a two-lens afocal system to fill the back aperture of the objective (Nikon, Plan Apo, ×100, numerical aperture 1.4, oil immersion). The setup was optimized to obtain a diffraction-limited focal spot. Measurements were typically carried out in a 50-μl solution dropped on the coverslip (buffer A + 50 mM NaCl). Fluorescence was collected by the same objective, separated from the excitation by a dichroic mirror (Chroma 700DCSPXR) and focused onto an avalanche photodiode (SPCM-AQR-14; PerkinElmer Life Sciences). An additional filter was used to reject the residual excitation light (Chroma 2P-Emitter E700SP). The detector was con-

nected to a digital correlator (ALV 6000) that calculates the normalized correlation function  $g(\tau)$  of fluorescence fluctuations according to Equation 1,

$$g(\tau) = \frac{\langle I(t) \cdot I(t + \tau) \rangle}{\langle I(t) \rangle^2} \quad (\text{Eq. 1})$$

where  $I(t)$  is the number of detected fluorescence photons per time unit. Assuming a three-dimensional Gaussian distribution of the excitation intensity, the FCS function for a free Brownian diffusion process is given by Equation 2,

$$g(\tau) = \frac{1}{N} \cdot \frac{1}{\left(1 + \frac{\tau}{\tau_D}\right) \cdot \sqrt{1 + \frac{\omega_0^2}{z_0^2} \cdot \frac{\tau}{\tau_D}}} \quad (\text{Eq. 2})$$

where  $N$  is the mean number of fluorophores in the excitation volume, and  $\tau_D$  is the translational diffusion time.  $\omega_0$  and  $z_0$  are the lateral and axial dimensions of the excitation volume, respectively. The mean translational diffusion time was determined by fitting autocorrelation curves using a Levenberg-Marquardt nonlinear least-squares fitting algorithm according to the analytical model (Equation 2). For two-photon excitation, the diffusion coefficient  $D$  is then calculated according to the equation,  $D = \omega^2/8\tau_D$ . The calibration of the excitation volume was performed using a 10 nM water solution of TAMRA ( $D = 2.8 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ ). The lateral  $\omega_0$  and axial  $z_0$  dimensions of the excitation volume were 0.45 and 1.5  $\mu\text{m}$ , respectively. The excitation power was adjusted using a neutral density filter. For our setup, we determined that an excitation power of 10 milliwatts was suitable for two-photon excitation of both TAMRA alone and IN<sup>T</sup> (fluorescence intensity exhibits a quadratic dependence, and  $\tau_D$  value is constant (absence of photobleaching) as a function of the incident power below 16 milliwatts). Recording times were typically 5 min (average of 10 cycles of 30 s).

**Time-resolved Fluorescence Anisotropy**—Fluorescence of Trp or Fl was used for hydrodynamic studies of free or DNA-bound PFV-1 IN, respectively (in buffer A + 50 mM NaCl). IN-DNA complexes were obtained using unlabeled PFV-1 IN and Fl-labeled 15- and 21-mer U5 PFV-1 ODN (Fl at the 5'-end of strand A). Correlation times ( $\theta$ ) were calculated from the two polarized fluorescence decays  $I_{\perp}(t)$  and  $I_{\parallel}(t)$ , using time-correlated single photon counting (21, 22). Briefly, the time scaling was 19.5 ps/channel, and 4096 channels were used. The excitation light pulse source was a Ti:Sapphire laser (Millennia-pumped Tsunami femtosecond laser; Spectra Physics) (repetition rate, 8 MHz) associated with a second or third harmonic generator tuned to 490 or 296 nm for Fl or Trp, respectively. The emission monochromator (ARC SpectraPro-150) was set to 530 or 340 nm ( $\Delta\lambda = 15 \text{ nm}$ ) for Fl or Trp, respectively. The two polarized components were collected alternately over a period of 30 s (total count of  $I_{\parallel}(t)$  was 15 millions, which is a condition compatible with the recovery of correlation times up to  $\sim 80 \text{ ns}$  using Trp fluorescence (7)).  $I_{\perp}(t)$  and  $I_{\parallel}(t)$  were analyzed by the maximum entropy method (23, 24) to determine the distributions of both lifetimes ( $\tau$ ) and rotational correlation times ( $\theta$ ), according to Equations 3 and 4.

$$I_{\parallel}(t) = \frac{1}{3} \times \left( \sum_{i=1}^n \alpha_i e^{-t/\tau_i} \left( 1 + 2 \sum_{j=1}^m \rho_j e^{-t/\theta_j} \right) \right) \quad (\text{Eq. 3})$$

$$I_{\perp}(t) = \frac{1}{3} \times \left( \sum_{i=1}^n \alpha_i e^{-t/\tau_i} \left( 1 - \sum_{j=1}^m \rho_j e^{-t/\theta_j} \right) \right) \quad (\text{Eq. 4})$$

with

$$\sum_{i=1}^n \alpha_i = 1 \quad (\text{Eq. 5})$$

and

$$\sum_{j=1}^m \rho_j = r'_0 \quad (\text{Eq. 6})$$

where  $\alpha_i$  represents the relative population characterized by the lifetime  $\tau_i$ , and  $\rho_j$  represents the initial anisotropy related to a motion characterized by the rotational correlation time  $\theta_j$ .  $r'_0$  is the apparent fundamental anisotropy value (typically 0.24–0.26 in the present study, which is below the fundamental anisotropy value, 0.3, for Trp at  $\lambda_{\text{ex}} = 296 \text{ nm}$ , suggesting that a fast component in the PFV-1 IN anisotropy decay cannot be resolved by the instrument as previously found for many proteins (25)). The decay of the total fluorescence intensity,  $I_T(t)$ , and the fluorescence anisotropy decay,  $r(t)$ , are then calculated from both polarized components according to Equations 7 and 8, respectively,

$$I_T(t) = I_{\parallel}(t) + 2G \times I_{\perp}(t) \quad (\text{Eq. 7})$$

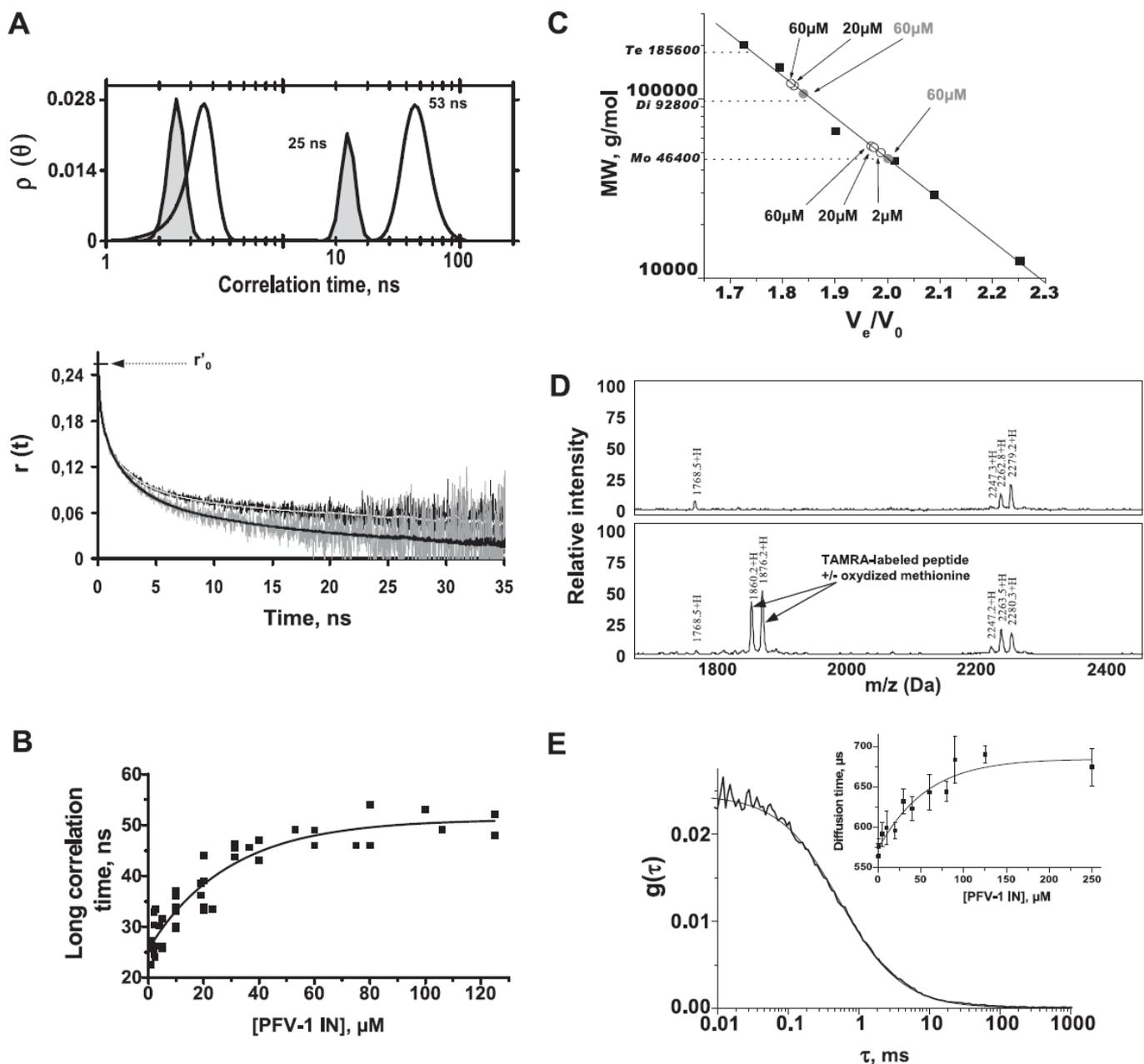
$$r(t) = \frac{I_{\parallel}(t) - G \times I_{\perp}(t)}{I_T(t)} = \sum_{j=1}^m \rho_j e^{-t/\theta_j} \quad (\text{Eq. 8})$$

where  $G$  represents the correction factor for the difference in the monochromator transmission between parallel and perpendicular polarized components.

## RESULTS

**TFA and FCS Reveal Distinct Solubility Properties between HIV-1 and PFV-1 INs**—TFA and FCS are complementary methods used to study the hydrodynamic properties of proteins, measuring the rotational and translational diffusion of molecules, respectively (25). Overall rotational diffusion or flexibility are major causes of light depolarization, and anisotropy decay analysis allows the determination of rotational correlation time ( $\theta$ ) distribution. Long correlation time ( $\theta_{\text{long}}$ ) represents the overall tumbling motion related to the protein hydrodynamic volume. In FCS, the temporal behavior of fluorescence fluctuations within a small excitation volume, described by the autocorrelation function, allows the determination of translational diffusion times.

PFV-1 IN was first studied by TFA using Trp fluorescence. Distributions of  $\theta$  are shown for two IN concentrations (Fig. 1A).  $\theta_{\text{long}}$  values were  $\sim 25$  and 53 ns for IN concentrations of 2 and 125  $\mu\text{M}$ , respectively. Given a molecular mass of 46.4 kDa



**FIGURE 1. Rotational and translational dynamics of PFV-1 IN.** *A, top*, rotational correlation time ( $\theta$ ) distribution of IN (gray, 2  $\mu\text{M}$ ; white, 125  $\mu\text{M}$ ) at 20 °C. Fluorescence anisotropy decays were obtained by monitoring the intrinsic Trp fluorescence. One additional short correlation time ( $\sim 400$  ps) was found under both concentration conditions (not shown). *Bottom*, corresponding experimental fluorescence anisotropy decays (gray, 2  $\mu\text{M}$ ; black, 125  $\mu\text{M}$ ) with line fits (black, 2  $\mu\text{M}$ ; white, 125  $\mu\text{M}$ ) resulting from maximum entropy method analysis (see “Experimental Procedures” for details). The lifetime distributions of PFV-1 IN, recovered from intensity decays (not shown), were as follows:  $\tau_1 = 0.28 \pm 0.03$  ns (29%),  $\tau_2 = 1.09 \pm 0.09$  ns (28%),  $\tau_3 = 2.6 \pm 0.3$  ns (20%),  $\tau_4 = 6.65 \pm 0.25$  ns (23%) for 2  $\mu\text{M}$  and  $\tau_1 = 0.24 \pm 0.01$  ns (21%),  $\tau_2 = 0.94 \pm 0.1$  ns (31%),  $\tau_3 = 2.4 \pm 0.2$  ns (21%),  $\tau_4 = 6.6 \pm 0.1$  ns (27%) for 125  $\mu\text{M}$ . *B*, long correlation time as a function of [IN]. *C*, analysis of PFV-1 IN by size exclusion chromatography. Experiments were performed at room temperature using a Superdex 200 column. Three concentrations of IN were loaded on the column. Only the monomeric species were detected using 2  $\mu\text{M}$ . An additional peak, corresponding to the dimeric form, appeared for the two highest IN concentrations (20 and 60  $\mu\text{M}$ ). No other species were detected.  $V_0$  and  $V_e$  are the exclusion volume of the column and the elution volume of the protein, respectively. The calibration sample (black squares) is composed of globular proteins of known molecular mass (see “Experimental Procedures”). The circles indicate the positions of elution peaks for various samples of PFV-1 IN, in the presence of either 0.05 M (gray circles) or 0.5 M (white circles) NaCl. The concentration of the loaded sample is explicitly mentioned in the figure. *D*, SELDI analysis of trypsin digestion of the unlabeled tagged (top) and TAMRA-labeled (bottom) PFV-1 IN. The peptides obtained after proteolysis were analyzed using a SEND-ID protein chip. Spectra are magnified in the region of interest, corresponding to the expected mass for the modified C-terminal peptide (NELEPKPQQFM; molecular mass 1860.16 Da) for the TAMRA-labeled protein. The entire spectra did not show any other differences between the two profiles (data not shown). The TAMRA-labeled IN was characterized by two peaks (1860.2 and 1876.9 Da). The first one is consistent with the expected theoretical value for the TAMRA-labeled C-terminal peptide (i.e. a single fluorophore was found as expected on the C terminus-specific tag), and the second one ( $\Delta\text{MW} = 16.7$  Da) corresponds to the same peptide containing an oxidized methionine. *E*, characterization of TAMRA-labeled PFV-1 IN by FCS; autocorrelation analysis of a 100 nM IN<sup>T</sup> solution. *Inset*, average diffusion time as a function of the total IN concentration. The IN<sup>T</sup> concentration was constant (100 nM), and total IN concentration was varied using unlabeled IN.

TABLE 1

Comparison of long correlation times obtained for PFV-1 and HIV-1 INs

IN	Theoretical $\theta_{\text{long}}$ monomeric form	Experimental $\theta_{\text{long}}$ monomeric form	Experimental $\theta_{\text{long}}$ higher order oligomeric forms
	<i>ns</i>	<i>ns</i>	<i>ns</i>
PFV-1	21.5	22.5–30 <sup>b</sup>	47–53 <sup>c</sup>
HIV-1	15	18–20 <sup>d</sup>	70–90 + >100 <sup>e</sup>

<sup>a</sup> Assuming partial specific volume ( $v$ ) of 0.73 ml/g and average hydration of 0.4 g(H<sub>2</sub>O)/g(protein) (44, 45).  $\theta_{\text{long}}$  is related to the volume of the rotating unit ( $V$ ) by  $\theta = \eta V/kT = \eta Mv/kT$ , where  $\eta$  is the viscosity,  $k$  is the Boltzman constant,  $T$  is the temperature (K), and  $M$  is the molecular weight.

<sup>b</sup> In the 1–5  $\mu\text{M}$  range.

<sup>c</sup> Monodisperse solution of dimeric species above 100  $\mu\text{M}$ .

<sup>d</sup> At submicromolar concentrations in the presence of detergent (7).

<sup>e</sup> Mainly tetrameric ( $\theta = 70$ –90 ns) in the absence of detergent below 200 nM; aggregation ( $\theta > 100$  ns) occurs above 200 nM (8, 26).

for monomeric PFV-1 IN, these values are consistent with a monomeric (Mo) and a dimeric (Di) form, respectively (Table 1).  $\theta_{\text{long}}$  increased continuously from 25 to 53 ns as a function of protein concentration with the half-transition state occurring at 20  $\mu\text{M}$  (Fig. 1B). The intermediary  $\theta_{\text{long}}$  values represent average correlation times during the transition and the observed transition reasonably corresponds to a Mo-Di equilibrium. This result was qualitatively confirmed by size exclusion chromatography (Fig. 1C).

These results significantly differ from those of HIV-1 IN (Table 1). Under similar conditions (*i.e.* absence of detergent), HIV-1 IN was mainly tetrameric below 200 nM and strongly aggregated above 200 nM (8, 26). This critical concentration was only 10-fold higher with detergent (7). Monomeric HIV-1 IN (32 kDa) (characterized by  $\theta_{\text{long}} = 19$  ns (7)) was obtained only at low submicromolar concentrations under detergent conditions.

The TGase-mediated TAMRA labeling of HIV-1 and PFV-1 INs (IN<sup>T</sup>) was next performed to study their translational diffusion properties by FCS. This enzymatic labeling (27) occurs specifically on a Gln residue of the PKPQQFM peptide substrate (fused to the C-terminal IN end) and leads to IN being labeled at a single position (confirmed by mass spectrometry analysis; Fig. 1D). HIV-1 IN<sup>T</sup> did not display autocorrelation curves that were satisfactory, due to the presence of bright spikes that originate in the presence of aggregates, regardless of the concentration used (data not shown). The number/frequency of bright spikes was slightly lower but remained significant upon the addition of detergent, indicating that irreversible aggregates were formed during the purification/labeling procedure. Thus, HIV-1 IN<sup>T</sup> was not further used in the present study. By contrast, PFV-1 IN<sup>T</sup> resulted in satisfactory autocorrelation curves (Fig. 1E) (no bright spike was detected during the acquisition). The fit yielded a diffusion time ( $\tau_D$ ) of 565  $\mu\text{s}$  at submicromolar concentrations. The FCS acquisition was repeated using a constant concentration of IN<sup>T</sup> and varying the concentration of unlabeled IN up to 250  $\mu\text{M}$  to study  $\tau_D$  and its concentration dependence (Fig. 1E, inset).  $\tau_D$  significantly increased from 565 to 680  $\mu\text{s}$  in this concentration range (again, no bright spike was detected in the different mixtures of labeled/unlabeled PFV-1 INs). The  $\tau_D$  ratio was 1.2, close to the theoretical ratio (1.3) for a Mo-Di transition. The FCS half-

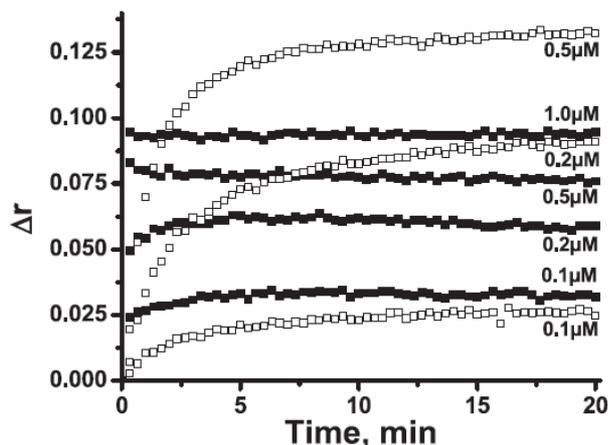


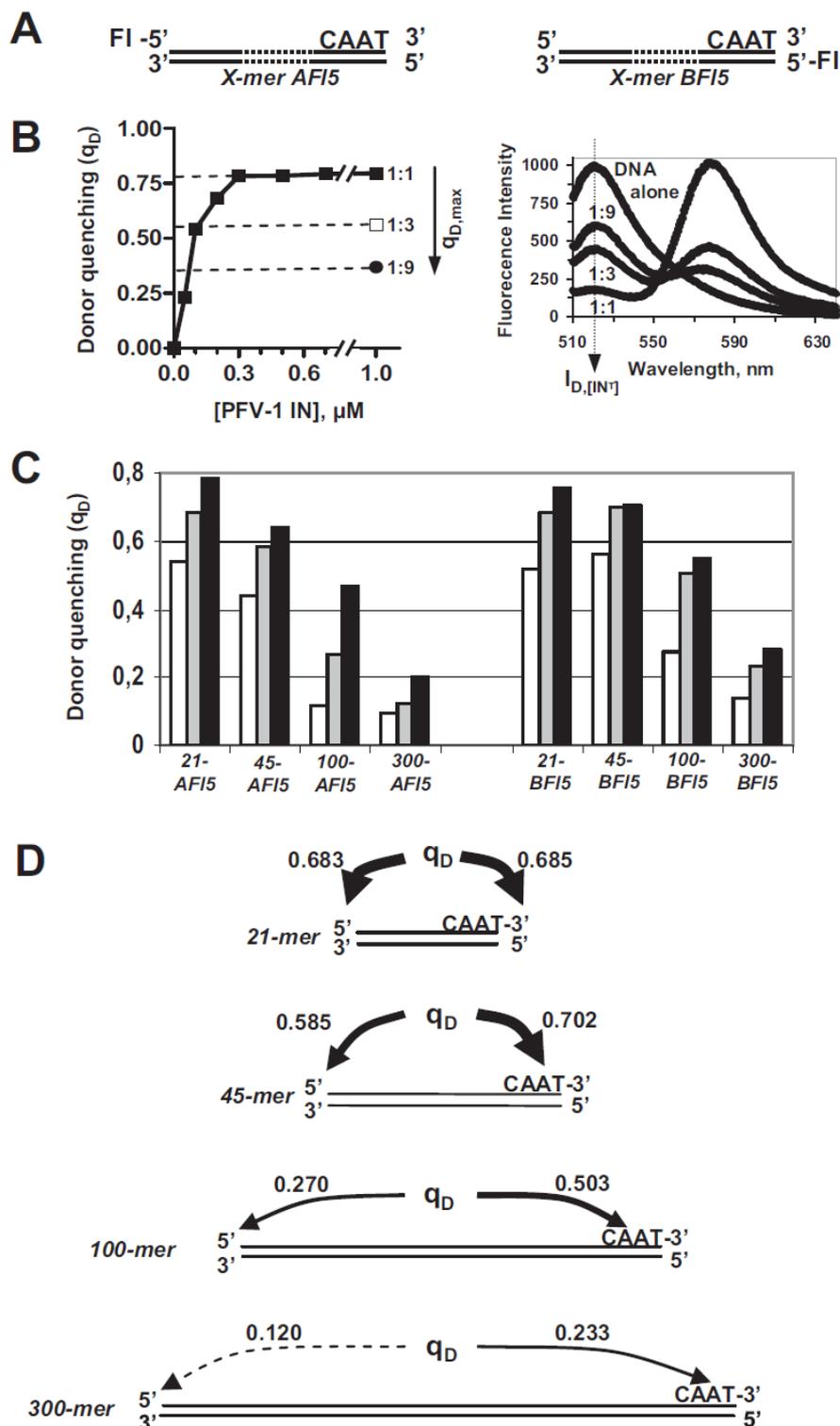
FIGURE 2. Differential DNA-binding kinetics of HIV-1 and PFV-1 INs. FI-labeled DNA that mimics either the HIV-1 or the PFV-1 U5 LTR extremity (double-stranded 21-mer, 4 nM) was mixed with HIV-1 (□) or PFV-1 IN (■), respectively, in buffer A + 50 mM NaCl. Protein concentrations are indicated in the figure. Steady-state fluorescence anisotropy values ( $r$ ) were measured at 25 °C as described under "Experimental Procedures."  $\Delta r(t) = r(t) - r_{\text{free DNA}}$ .

transition concentration (30  $\mu\text{M}$ ) was consistent with that observed in TFA.

We have previously shown that the HIV-1 IN DNA-binding step is very slow ( $k_{\text{on}} = 0.23 \text{ min}^{-1}$ ) (28). We suggested that these slow kinetics originate in a prebinding transition from higher order multimeric states (tetramers and aggregates) to smaller molecular species (Mo), whereas the DNA-bound dimeric form correlates with optimal 3'-P activity (21, 26, 28). Accordingly, Faure *et al.* (29) found that the Mo·DNA precedes the Di·DNA complexes by analyzing the time-dependent formation of LTR-cross-linked species. Interestingly, fluorescence anisotropy shows that a rapid equilibrium is reached after mixing PFV-1 IN and DNA, as compared with the much slower formation kinetics of HIV-1 IN·DNA complexes (Fig. 2), confirming TFA and FCS data (lack of detection of large PFV-1 IN molecular species before DNA binding). Therefore, the conversion of higher order multimers to Mo, responsible for the slow DNA-binding step, is a characteristic of HIV-1 IN. PFV-1, which is mainly monomeric at low micromolar concentrations, is characterized by a rapid DNA-binding step.

*Study of PFV-1 IN<sup>T</sup>-DNA Interaction by FRET Reveals a Preferential DNA Binding for the 3'-End Processing Site*—FRET experiments were performed to get deeper insight into the positioning of IN on the DNA substrate. DNA of various sizes (increasing from 21- to 300-mer) were FI-labeled at the 5'-end, either at the processing end or at the opposite end (Fig. 3A). Fig. 3B shows one example for the quenching donor ( $q_D$ ) plotted against PFV-1 IN<sup>T</sup> concentration. Only the quenching donor was measured (at 520 nm), not the acceptor sensitization (at 580 nm), since the acceptor concentration was a varying parameter along the titration.  $q_D$  increased as [IN<sup>T</sup>] increased and reached a plateau corresponding to a maximum quenching efficiency  $q_{D,\text{max}}$  of 78% (the donor emission did not change using equivalent amounts of unlabeled IN).  $q_{D,\text{max}} = 78\%$  is consistent with the theoretical value (at least 75%), assuming a labeling yield of 50% (see "Experimental Procedures") and a minimal IN/DNA stoichiometry of 2:1 (see below). The  $q_{D,\text{max}}$  value was decreased continuously by decreasing the IN<sup>T</sup>/unla-

## Heterogeneity and Catalytic Properties of IN-DNA Complexes



**FIGURE 3.** FRET between DNA substrates and PFV-1 IN. FRET between FI-labeled DNA substrates (donor) of various lengths and IN<sup>T</sup> (acceptor) was measured as a function of [IN<sup>T</sup>], as described under “Experimental Procedures.” **A**, the different DNA substrates used in the FRET study. The FI donor was attached at the 5′-end of either the A or B strand, giving *X-mer AFI5* or *BFI5*, respectively ( $X = 21, 45, 100$ , or  $300$ ). **B**, quenching donor ( $q_D$ ) between 21-mer *AFI5* and IN<sup>T</sup> (measured at 520 nm). The plateau value depends on the labeled/unlabeled IN ratio (indicated on the plot). **Right**, emission spectra for the various mixtures of labeled and unlabeled IN (total concentration,  $1 \mu\text{M}$ ) in the presence of  $20 \text{ nM}$  21-mer *AFI5* and IN<sup>T</sup>. **C**, comparison of  $q_D$  values between the different DNA substrates. [IN<sup>T</sup>] =  $100$  (white squares),  $200$  (gray squares), or  $300 \text{ nM}$  (black squares). **D**, summary of the differential quenching amplitude between the two DNA ends as a function of the DNA size for  $200 \text{ nM}$  IN<sup>T</sup>. All FRET experiments were performed in  $20 \text{ mM}$  Hepes, pH 7.5, containing  $1 \text{ mM}$  dithiothreitol,  $10 \text{ mM}$  MgCl<sub>2</sub>, and  $50 \text{ mM}$  NaCl.

beled IN ratio (Fig. 3B), confirming that the decrease in the donor intensity was actually due to resonance energy transfer.

We found that  $q_D$  was dependent on the DNA length (Fig. 3C); for a given IN concentration,  $q_D$  was systematically higher for short DNAs. This result was obtained although the apparent affinity decreases as the DNA size decreases (Table 2), demonstrating that the IN positioning on DNA was the main factor influencing the DNA size dependence of  $q_D$ . Only positions of IN<sup>T</sup> in close proximity to donor ( $<60\text{--}80 \text{ \AA}$  for the FI-TAMRA pair) can significantly influence the energy transfer. The continuous decrease in the  $q_D$  value obtained with DNA molecules of greater sizes demonstrates that IN may occupy many internal positions in DNA, leading to a “pearl chain” structure. Such a polymerization on long DNAs is consistent with the observation of light scattering that occurs for only 100- and 300-mers at IN concentrations above  $600 \text{ nM}$  (data not shown).

Interestingly, for 45-, 100-, and 300-mer DNA,  $q_D$  was systematically higher (for a given DNA size) when the FI donor was attached on the terminal processing site, up to 2-fold higher for 300-mer (FRET results are summarized in Fig. 3D). No such difference was observed with the 21-mer. Our results indicate a significant preference for IN binding at the processing end. This is consistent with the absence of bias for the short 21-mer substrates, where a difference in the quenching donor between the two ends is not expected, since the dimension of DNA is comparable with both the overall dimension of IN and the Förster distance.

*The Anisotropy-based Assay Confirms a Specific DNA-binding Mode for PFV-1 but Not for HIV-1 IN*—We next assessed the specific DNA-binding of PFV-1 IN (suggested by FRET) by competition experiments. Unlabeled PFV-1 IN was incubated with FI-labeled DNA, 45-mer PFV-1 U5 (Fig. 4A), or 45-mer random

TABLE 2

Apparent  $K_d$  values of PFV-1 IN as a function of DNA size and ionic strength

Fl-labeled ODNs of various sizes mimicking the cognate U5 DNA end (12 nm) were mixed with increasing concentrations of unlabeled PFV-1 IN in the presence of 10 mM  $Mg^{2+}$ . DNA binding was measured by recording the steady-state anisotropy value as described under "Experimental Procedures."

DNA size		[NaCl]	$K_{d,app}$
bp	mm	$\mu M$	
15	50	0.56	
	150	3.6	
	200	ND <sup>a</sup>	
21	50	0.13	
	150	2.1	
	200	3.5	
45	50	0.1	
	150	0.4	
	200	1.6	
65	50	0.06	
	150	0.39	
	200	0.97	
100	50	0.05	
	150	0.29	
	200	ND	
300	50	0.06	
	150	0.37	
	200	ND	

<sup>a</sup> ND, not determined.

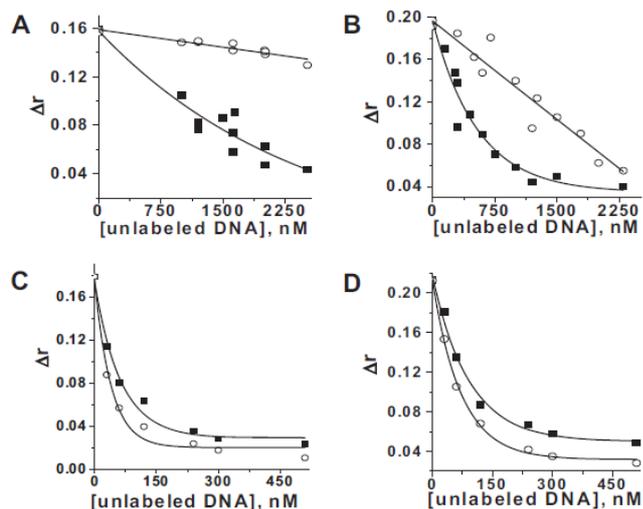


FIGURE 4. Comparison of HIV-1 and PFV-1 IN DNA binding properties. IN (A and B, PFV-1; C and D, HIV-1) was incubated with 4 nM Fl-labeled 45-mer DNA (A and C, cognate sequence; B and D, random sequence) and varying concentrations of unlabeled 45-mer DNA competitor (■, cognate sequence; ○, random sequence) in buffer A + 50 mM NaCl (See "Experimental Procedures" for ODN sequences). Fluorescence anisotropy was measured as described under "Experimental Procedures."  $\Delta r = r_{\text{complex}} - r_{\text{free DNA}}$ . [IN] = 700 nM (PFV-1) or 400 nM (HIV-1).

sequence (Fig. 4B) in the presence of increasing concentrations of unlabeled DNA (45-mer PFV-1 U5 or random sequence). The decrease in the anisotropy value was then recorded. Using the cognate Fl-labeled DNA, unlabeled cognate and random sequences competed with different  $IC_{50}$  values (1.4 and >2.5  $\mu M$ , respectively) (Fig. 4A). Consistently, the competition was easier when the Fl-labeled DNA was random, with corresponding values of 0.5 and 1.6  $\mu M$ , respectively (Fig. 4B). These results indicate that PFV-1 IN has a higher apparent affinity for the cognate sequence. The  $IC_{50}$  value found for the PFV-1 IN-45-mer U5 complex (1.4  $\mu M$  at 50 mM NaCl) was used to estimate the apparent  $K_d$  value ( $K_{d,app}$ ) according to Refs. 30 and 31) (the

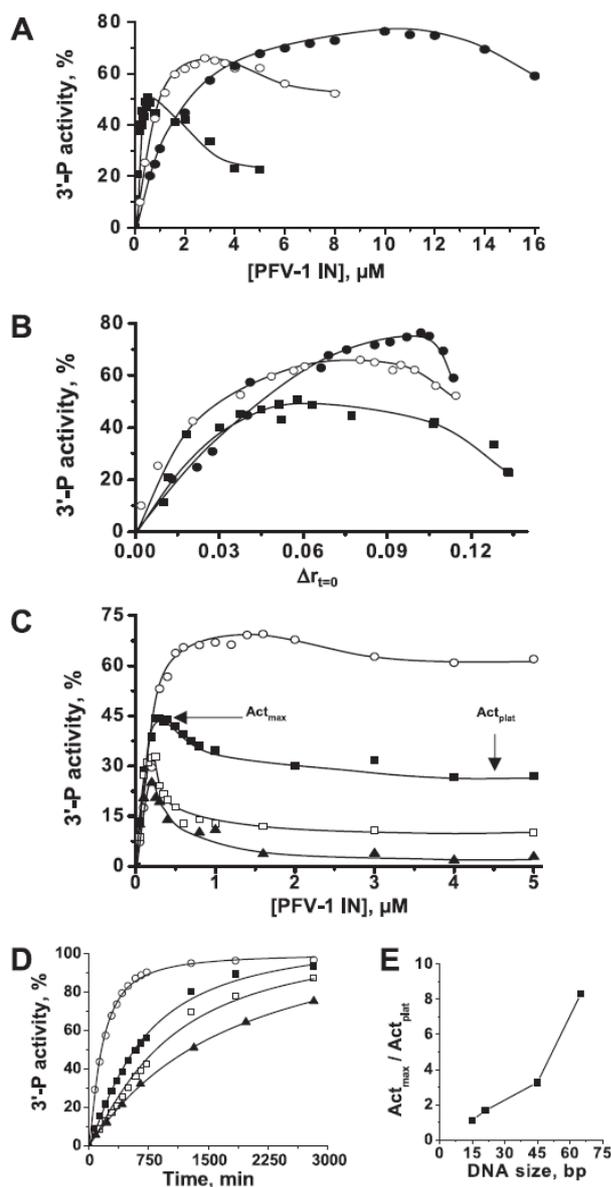
Cheng-Prusoff relationship was not used here because [IN] > [DNA substrate]). The calculated  $K_{d,app}$  from the competition assay (0.13  $\mu M$ ) was then compatible with the corresponding  $K_{d,app}$  obtained by direct titration measurement (0.1  $\mu M$  in Table 2), indicating no influence or only a slight influence of the Fl moiety on the DNA binding properties of IN.

Preferential/specific DNA-binding of HIV-1 IN is generally difficult to assess *in vitro* (10, 12, 32) despite the fact that (i) specific and nonspecific complexes are distinct in term of radius of gyration (33), and (ii) several residues (Gln<sup>148</sup>, Tyr<sup>143</sup>, Lys<sup>156</sup>, Lys<sup>159</sup>, Lys<sup>14</sup>) appear to be involved in such specific contacts before and/or after the 3'-P reaction (11, 32, 34–36). In particular, it was recently shown that Gln<sup>148</sup> is also involved in maintaining the complex stability for the subsequent strand transfer reaction by specifically contacting the 5'-LTR end of the donor (viral) DNA (34, 35). Competition experiments were repeated with HIV-1 IN and confirm that no significant difference can be observed between the cognate and a random 45-mer DNA sequence, irrespective of the Fl-labeled DNA sequence (Fig. 4, C and D). Moreover, HIV-1 and PFV-1 INs display different apparent affinities for their respective DNA substrate (Fig. 4, compare A and B with C and D). For 45-mer DNA, the specificity and the absence of specificity for PFV-1 and HIV-1, respectively, were obtained regardless of the ionic strength (50 or 150 mM NaCl). For 21-mer DNA, PFV-1 IN displayed differential DNA binding between cognate and random sequences at 150 mM NaCl (two different random sequences were tested and gave similar results), whereas no specific DNA binding was observed at 50 mM NaCl (data not shown). Again, no specificity was observed with HIV-1 IN using 21-mer DNA at either salt concentration (data not shown).

**Comparative Study of Catalytic Properties of HIV-1 and PFV-1 INs**—Taking into account that solubility and DNA binding properties of HIV-1 and PFV-1 INs were significantly different, we next compared catalytic features of both enzymes using a fluorescence 3'-P assay (21). Briefly, this assay is based on steady-state anisotropy ( $r$ ) and allows the separation of DNA-binding and catalytic parameters of 3'-P reaction as well as the study of real time kinetics. A simultaneous quantitative analysis of DNA binding and dinucleotide release is possible, since both steps strongly influence the molecular size of the fluorescent moiety when Fl is linked to the released dinucleotide.

For HIV-1, we have previously found that the  $r$  value, as obtained after DNA-binding and before catalysis ( $r_{t=0}$ ), was fully predictive of the subsequent activity at low IN/DNA ratios, according to the fractional saturation function (21). By contrast, for higher ratios,  $r$  continued to increase but the 3'-P activity significantly decreased. The variable  $r$  is dependent on both the fractional DNA saturation and the molecular size of the complex. Therefore, higher order IN multimers/aggregates are detrimental to 3'-P activity and account for the characteristic bell-shaped curves when plotting activity *versus* IN concentration or the  $r_{t=0}$  parameter. We obtained a similar behavior with PFV-1 IN using the cognate 21-mer PFV-CAAT substrate when the 3'-P activity was plotted against IN concentration (Fig. 5A) or  $r_{t=0}$  (Fig. 5B). This suggests that, although HIV-1 and PFV-1 INs, free in solution, are characterized by

## Heterogeneity and Catalytic Properties of IN-DNA Complexes



**FIGURE 5. Influences of DNA size and ionic strength on the PFV-1 IN 3'-P activity.** A and B, ionic strength effect. 3'-P activity was determined using the 21-mer DNA substrate as a function of [IN] (A) or the  $\Delta r_{t=0}$  value (B). [NaCl] = 50 mM (■), 150 mM (○), or 200 mM (●).  $\Delta r_{t=0} = r_{t=0} - r_{\text{free DNA}}$ .  $r_{t=0}$  represents the  $r$  value after the DNA-binding step (before the start of the reaction). C, DNA size effect. The response of the 3'-P activity to IN concentration was studied with 15-mer (○), 21-mer (■), 45-mer (□), and 65-mer (▲), mimicking the U5 LTR extremity. [NaCl] = 50 mM. Activities reported on the y axis (A–C) correspond to a 420-min incubation with 12 nM FI-labeled DNA at 37 °C and were calculated as described under “Experimental Procedures.” D, time course of 3'-P for 15–65-mer (symbols are as in C). Experimental conditions correspond to optimal conditions in C for each DNA size. E,  $\text{Act}_{\text{max}}/\text{Act}_{\text{plat}}$  ratio as a function of the DNA length.  $\text{Act}_{\text{max}}$  and  $\text{Act}_{\text{plat}}$  are defined under “Results” and in C.

distinct self-association states, they display comparable enzymatic features for the 3'-P catalytic reaction. HIV-1 IN was previously shown to work very slowly as a single-turnover enzyme for 3'-P (rate constant =  $0.004 \text{ min}^{-1}$ ) (28). Here, PFV-1 IN possesses similar slow single-turnover kinetics (Fig. 5D), suggesting an intrinsic single-turnover property of retroviral INs that is clearly independent of the initial aggregation

**TABLE 3**

### Long correlation times of PFV-1 IN-DNA complexes

TFA experiments were performed with 12 nM DNA in buffer A + 50 mM NaCl at 20 °C, as described under “Experimental Procedures.”

DNA size <i>bp</i>	[PFV-1 IN] <i>nM</i>	$\theta_{\text{long}}^a$
		<i>ns</i>
15	1200	$63 \pm 6$
21	400	$53 \pm 8$
21	600	$61 \pm 5$
21	1200	$>100$

<sup>a</sup> The correlation time distributions of free ODNs as a function of the DNA size were previously published (21).  $\theta_{\text{long}} = 6$  and 9.5 ns at 20 °C for free ODN<sub>15-mer</sub> and free ODN<sub>21-mer</sub>, respectively.

state. Interestingly, the 3'-P activity at the top of the bell-shaped curve ( $\text{Act}_{\text{max}}$ ) was significantly better (from 50 to 75%) with greater ionic strength (from 50 to 200 mM NaCl) (Fig. 5A), although more IN was required to obtain optimal activity, because the overall IN-DNA affinity is weakened by ionic strength (Table 2). This result suggests that large complexes, responsible for the activity drop off, were more sensitive to ionic strength than catalytically competent complexes.

Furthermore, the 3'-P activity is better at higher ionic strength for a given  $r_{t=0}$  value in the decreasing phase (Fig. 5B), confirming that  $r_{t=0}$  accounts for both the number of 3'-P-competent complexes and the less active large complexes. The enhanced activity in high salt concentration at the optimal  $r_{t=0}$  value confirms that ionic strength differentially affects the two types of complexes and delays the formation of large complexes.

We obtained bell-shaped curves for all DNA lengths tested (Fig. 5C). However, we found that the optimal activity ( $\text{Act}_{\text{max}}$ ) was strongly length-dependent, and better 3'-P activity was obtained using shorter ODNs (3-fold higher for 15-mer than that of 65-mer) (confirmed by complete kinetics study in Fig. 5D). Again,  $\text{Act}_{\text{max}}$  did not occur at similar IN concentrations. The optimal concentration depends on DNA size, and this is consistent with apparent  $K_d$  values showing a better IN affinity for long DNAs (Table 2). Interestingly, the activities in the decreasing phase reached a plateau ( $\text{Act}_{\text{plat}}$ ) (Fig. 5C). The decreasing phase was more pronounced for longer DNA substrates, as evidenced when plotting the  $\text{Act}_{\text{max}}/\text{Act}_{\text{plat}}$  ratio versus DNA size (Fig. 5E). These results indicate that shortening the DNA substrate improves the number of catalytically competent complexes over the number of less active large complexes. The stimulating effect of ionic strength on  $\text{Act}_{\text{max}}$  (as described in Fig. 5A for the 21-mer) was reproduced with the 15-mer DNA substrate ( $\text{Act}_{\text{max}} = 90\%$  at 200 mM NaCl; data not shown). Surprisingly, no such effect was observed for the longer 45- and 65-mer DNAs (data not shown). This result will be further discussed below.

Fluorescence of FI was next used for the TFA study of the hydrodynamics of PFV-1 IN-DNA complexes using 15- and 21-mer FI-labeled ODNs (Table 3). At the optimal IN concentration for activity ( $\text{Act}_{\text{max}}$ ),  $\theta_{\text{long}}$  values were consistent with dimeric forms of PFV-1 IN bound to short DNAs (53–63 ns). Above this concentration, we detected large complexes ( $\theta_{\text{long}} > 100 \text{ ns}$ ), consistent with the bell-shaped curves shown in Fig. 5A. Therefore, the dimeric form appears to be the most catalytically competent form for 3'-P, and higher order multimers or

aggregates, obtained for high IN/DNA ratio, led to suboptimal activity. These results strongly parallel those previously obtained with HIV-1 (21).

## DISCUSSION

We used various biophysical techniques to address the problem of IN solubility and found that HIV-1 and PFV-1 INs are characterized by different solubility properties. PFV-1 IN, free in solution, was found to be more soluble than HIV-1 IN. Moreover, fluorescence anisotropy and FRET experiments highlight sequence-specific IN-DNA recognition that is measurable only with PFV-1 IN. However, most of the enzymatic features are similar for both types of INs (slow single turnover kinetics, dimeric form responsible for optimal 3'-P). In addition, although more soluble, PFV-1 IN can lead to higher order multimers or aggregates when bound to its DNA substrate under conditions of high IN/DNA ratio (leading to suboptimal activity), as previously found for HIV-1 IN (21). Finally, DNA size and ionic strength were found to be important parameters that modulate the number of higher order complexes. In addition, we have demonstrated the feasibility of the enzymatic labeling of a retroviral integrase with a low molecular weight fluorophore. In contrast to conventional chemical labeling methods, such a strategy leads to a site-specific labeling of the protein of interest, avoiding multiply labeled proteins and heterogeneous samples, and then is particularly well suited for FRET and FCS studies. Regrettably, bright spikes were detected by FCS with HIV-1 IN<sup>T</sup>, suggesting the presence of aggregates, whereas PFV-1 IN<sup>T</sup> resulted in satisfactory autocorrelation curves with no bright spike detected, reinforcing the idea that HIV-1 IN has a higher propensity for aggregation than PFV-1 IN.

In previous studies, HIV-1 IN was found mainly tetrameric, below 200 nm in the absence of detergent, and significant protein aggregation was observed above 200 nm (8, 26). In the presence of detergent, this critical concentration was found in the low micromolar range. Here, PFV-1 IN was found mainly monomeric in the low micromolar range, in the absence of any detergent and DNA. Using PFV-1 IN, either unlabeled (in size exclusion chromatography and TFA experiments) or TGase-mediated TAMRA-labeled (in FCS experiments), we have clearly identified a Mo  $\leftrightarrow$  Di equilibrium for the protein, free in solution. TFA and FCS gave approximately the same half-transition concentration, 20–30  $\mu$ M. PFV-1 IN is therefore characterized by a Mo  $\leftrightarrow$  Di equilibrium, and aggregation does not significantly occur in the absence of DNA during the transition. The reason for better solubility than HIV-1 IN remains unclear, since no evident differences appear when comparing the global hydrophobic profiles of both INs (not shown). Mutating only a few surface residues significantly improves HIV-1 IN solubility. For instance, the single mutation F185K improves the solubility of the entire protein (6). It is then possible that only a small number of key residues may account for enhanced PFV-1 IN solubility. Due to the poor sequence similarity between both INs (~15%), a greater understanding of solubility determinants requires further structural studies.

Interestingly, using the TAMRA-labeled PFV-1 IN and fluorescein-labeled DNA substrates, we found that FRET was systematically more efficient when fluorescein was attached at the

processing as compared with the nonprocessing end. This difference was observed only with long DNA substrates (from 45- to 300-mer) and not with a short 21-mer DNA substrate, with size comparable with the Förster distance. This result highlights a preference of DNA binding for the processing end and was confirmed by competitive DNA-binding experiments. Altogether, these results suggest that, using PFV-1 IN, it is possible to distinguish between specific and nonspecific complexes, not only at the catalytic level (as typically found for HIV-1 (11, 21)) but also at the DNA binding level. In our opinion, these results do not reveal a differential DNA binding property of both INs but rather, most likely, that the aggregative properties and high propensity of HIV-1 IN to establish nonspecific contacts mask *in vitro* an intrinsic ability of the protein to bind preferentially to its cognate sequence. However, FRET and competition experiments suggest that the specificity, although significant, is modest for PFV-1 IN. This could be due to the inherent nonspecific DNA-binding mode (responsible for the binding to the target DNA) that can be mainly ascribed to the C-terminal domain, as for the HIV-1 protein (12), although we cannot exclude the possibility that this property is also mediated by the N-terminal region in the case of PFV-1 IN (18). Furthermore, the catalytic domain, which ensures 3'-P and the subsequent joining reaction into a large variety of target DNA sequences, should be also in part responsible for nonspecific DNA-binding.

FRET analysis, as a function of DNA size, also suggests a nonspecific DNA binding mode at internal positions onto long DNA substrates. In addition, we found that the 3'-P catalytic process was stimulated by (i) shortening the DNA size and (ii) increasing the ionic strength. Nevertheless, the stimulating effect of ionic strength was not observed for long DNA substrates such as 45- and 65-mer. This result, together with the FRET data, suggests the existence of two types of higher order complexes. IN polymerization, mediated by nonspecific internal IN-DNA interactions, is the main phenomenon occurring on long DNAs and is strongly detrimental to 3'-P activity. Consistently, the Act<sub>max</sub>/Act<sub>plat</sub> ratio (as defined in Fig. 5) is lower for 15- and 21-mer DNAs, since no polymerization is expected to occur on short DNAs for evident steric reasons. However, their slight decreasing phases may be assigned to the presence of aggregates, mediated by nonspecific IN-IN interactions at high IN/DNA ratios. Only a differential effect of high salt concentration on nonspecific IN-DNA and IN-IN interactions may explain why ionic strength only stimulates Act<sub>max</sub> of short DNAs. Ionic strength disrupts more efficiently nonspecific IN-IN interactions than nonspecific IN-DNA interactions. This also explains why the specific interaction between PFV-1 IN and the short 21-mer DNA was only evidenced under high salt conditions (see above), since aggregates on DNA underestimate detection of specific complexes at low ionic strength. Altogether, our results show that two types of large complexes are present at high IN/DNA ratios, and both are detrimental for IN activity. The polymerization process is mainly influenced by the DNA length, whereas aggregation is more sensitive to the ionic strength. Consequently, an additive/synergic positive effect on the IN activity is observed by combining both param-

## Heterogeneity and Catalytic Properties of IN·DNA Complexes

eters (*i.e.* shortening the DNA length and increasing the ionic strength).

FRET analysis indicates that IN on long DNAs can be positioned at internal DNA sites due to its nonspecific DNA-binding mode, decreasing the 3'-P efficiency. The DNA length dependence of 3'-P activity raises the question of the particularly low 3'-P efficiency, which can be expected for long DNAs, such as physiological viral DNAs (about 9 and 13 kbp long for HIV-1 and PFV-1, respectively). To date, the mechanism by which IN is specifically positioned at the processing site *in vivo* is still unclear. Although IN prefers to bind to its cognate DNA sequence above other sequences, the relative specificity of IN described in this study may not be sufficient to ensure its selective *in vivo* positioning at the processing ends. It has been previously proposed for HIV-1 that the nucleocapsid protein binds tightly to viral DNA, except at the DNA ends (37). Additionally, this protein stimulates IN activity *in vitro* (38) and may indirectly assist the correct positioning of IN onto the processing ends. It is important to note that a nucleocapsid, characterized by a typical two canonical cysteine-histidine motifs, is not present in PFV-1, although the PFV-1 C terminus region of Gag contains three glycine-arginine motifs called "GR boxes," which are supposed to display similar properties (39, 40).

The PFV-1 IN·DNA complexes, as obtained under optimal conditions for 3'-P activity ( $\text{Act}_{\text{max}}$  in Fig. 5), were further studied by time-resolved fluorescence anisotropy. Using short DNA substrates (15- and 21-mer), long rotational correlation times were found to be compatible with dimeric forms (Table 3), suggesting that Di is the most catalytically competent form for 3'-P, as was previously found for HIV-1 IN (21, 29), whereas large complexes are characterized by suboptimal 3'-P activity. DNA primarily promotes IN dimerization, since Di are present at concentrations in which free PFV-1 IN is monomeric (low micromolar range). Most likely, a tetramer (Di of Di) is required for integration, as shown for HIV-1 when two LTR ends are simultaneously present in the context of the synaptic complex (41) or when IN is bound to a three-way junction DNA substrate mimicking an integration intermediate (42). The ability of the tetrameric species to perform 3'-P is still controversial (29, 43), and the present data cannot clearly address the proper 3'-P activity of the Te, since the decreasing phase is certainly related to an heterogeneous mixture of large complexes, including tetrameric forms. Recently, a kinetic study of ASV  $\text{Mn}^{2+}$ -dependent 3'-P suggests that a  $\text{Di} \leftrightarrow \text{Te}$  equilibrium accounts for the biphasic behavior of the single-turnover 3'-P, with Te responsible for the fast phase (43). We found a monophasic behavior for both PFV-1 (this study; Fig. 5D) and HIV-1 (21)  $\text{Mg}^{2+}$ -bound INs, with rate constants compatible with the slow ASV phase. Apparent discrepancies between HIV-1, PFV-1, and ASV may reflect intrinsic or metal-dependent differences for  $\text{Di} \leftrightarrow \text{Te}$  or/and  $\text{Te} \leftrightarrow \text{aggregates}$  transitions onto DNA. However, HIV-1 and PFV-1 INs, which display similar monophasic behaviors, are characterized by different single-turnover rate constants (for a given DNA size, *i.e.* 21-mer DNA substrate at 50 mM NaCl,  $k_{\text{HIV-1}} = 0.004 \text{ min}^{-1}$  (21) and  $k_{\text{PFV-1}} = 0.0028 \text{ min}^{-1}$  (this study)). Apparent affinities for DNA substrate were also significantly different ( $K_{d,\text{app}} = 40$  and 130 nM for HIV-1 (21) and PFV-1 (Table 2), respectively, for

21-mer DNA substrate at 50 mM NaCl). Although most of the enzymatic properties described here are qualitatively similar between HIV-1 and PFV-1 INs, the two enzymes display substantial quantitative differences; PFV-1 IN has lower affinity for the DNA substrate and lower catalytic efficiency as compared with HIV-1 IN; the reasons for these differences remain to be elucidated. Although the sequence homology between HIV-1 and PFV-1 INs is low, our data suggest that both proteins, which display similar enzymatic properties, at least qualitatively, have structural homology. Considering its solubility characteristics, PFV-1 IN could therefore represent a good model for further structural studies of retroviral INs.

## REFERENCES

1. Sinha, S., and Grandgenett, D. P. (2005) *J. Virol.* 79, 8208–8216
2. Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* 97, 8233–8238
3. Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001) *EMBO J.* 20, 7333–7343
4. Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R., and Davies, D. R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 9150–9154
5. Cherepanov, P., Ambrosio, A. L., Rahman, S., Ellenberger, T., and Engelman, A. (2005) *Proc. Natl. Acad. Sci. U. S. A.* 102, 17308–17313
6. Jenkins, T. M., Engelman, A., Ghirlando, R., and Craigie, R. (1996) *J. Biol. Chem.* 271, 7712–7718
7. Deprez, E., Tauc, P., Leh, H., Mouscadet, J. F., Auclair, C., and Brochon, J. C. (2000) *Biochemistry* 39, 9275–9284
8. Leh, H., Brodin, P., Bischerour, J., Deprez, E., Tauc, P., Brochon, J. C., LeCam, E., Coulaud, D., Auclair, C., and Mouscadet, J. F. (2000) *Biochemistry* 39, 9285–9294
9. Marchand, C., Johnson, A. A., Karki, R. G., Pais, G. C., Zhang, X., Cowanage, K., Patel, T. A., Nicklaus, M. C., Burke, T. R., Jr., and Pommier, Y. (2003) *Mol. Pharmacol.* 64, 600–609
10. Agapkina, J., Smolov, M., Barbe, S., Zubin, E., Zatssep, T., Deprez, E., Le Bret, M., Mouscadet, J. F., and Gottikh, M. (2006) *J. Biol. Chem.* 281, 11530–11540
11. Esposito, D., and Craigie, R. (1998) *EMBO J.* 17, 5832–5843
12. Engelman, A., Hickman, A. B., and Craigie, R. (1994) *J. Virol.* 68, 5911–5917
13. Moebes, A., Enssle, J., Bieniasz, P. D., Heinkelein, M., Lindemann, D., Bock, M., McClure, M. O., and Rethwilm, A. (1997) *J. Virol.* 71, 7305–7311
14. Yu, S. F., Sullivan, M. D., and Linnal, M. L. (1999) *J. Virol.* 73, 1565–1572
15. Cherepanov, P. (2007) *Nucleic Acids Res.* 35, 113–124
16. Enssle, J., Moebes, A., Heinkelein, M., Panhuysen, M., Mauer, B., Schweizer, M., Neumann-Haefelin, D., and Rethwilm, A. (1999) *J. Gen. Virol.* 80, 1445–1452
17. Pahl, A., and Flugel, R. M. (1995) *J. Biol. Chem.* 270, 2957–2966
18. Lee, H. S., Kang, S. Y., and Shin, C. G. (2005) *Mol. Cells* 19, 246–255
19. Oh, Y. T., and Shin, C. G. (1999) *Biochem. Mol. Biol. Int.* 47, 621–629
20. Delelis, O., Petit, C., Leh, H., Mbemba, G., Mouscadet, J. F., and Sonigo, P. (2005) *Retrovirology* 2, 31–49
21. Guiot, E., Carayon, K., Delelis, O., Simon, F., Tauc, P., Zubin, E., Gottikh, M., Mouscadet, J. F., Brochon, J. C., and Deprez, E. (2006) *J. Biol. Chem.* 281, 22707–22719
22. Laboulais, C., Deprez, E., Leh, H., Mouscadet, J. F., Brochon, J. C., and Le Bret, M. (2001) *Biophys. J.* 81, 473–489
23. Brochon, J. C. (1994) *Methods Enzymol.* 240, 262–311
24. Livesey, A. K., and Brochon, J. C. (1987) *Biophys. J.* 52, 693–706
25. Lakowicz, J. (2006) in *Principles of Fluorescence Spectroscopy*, 3rd Ed., Springer-Verlag New York Inc., New York
26. Deprez, E., Tauc, P., Leh, H., Mouscadet, J. F., Auclair, C., Hawkins, M. E., and Brochon, J. C. (2001) *Proc. Natl. Acad. Sci. U. S. A.* 98, 10090–10095
27. Taki, M., Shiota, M., and Taira, K. (2004) *Protein Eng. Des. Sel.* 17, 119–126
28. Smolov, M., Gottikh, M., Tashlitskii, V., Korolev, S., Demidyuk, I., Bro-

- chon, J. C., Mouscadet, J. F., and Deprez, E. (2006) *FEBS J.* 273, 1137–1151
29. Faure, A., Calmels, C., Desjobert, C., Castroviejo, M., Caumont-Sarcos, A., Tarrago-Litvak, L., Litvak, S., and Parissi, V. (2005) *Nucleic Acids Res.* 33, 977–986
  30. Linden, J. (1982) *J. Cyclic Nucleotide Res.* 8, 163–172
  31. Deprez, E., Barbe, S., Kolaski, M., Leh, H., Zouhiri, F., Auclair, C., Brochon, J. C., Le Bret, M., and Mouscadet, J. F. (2004) *Mol. Pharmacol.* 65, 85–98
  32. Jenkins, T. M., Esposito, D., Engelman, A., and Craigie, R. (1997) *EMBO J.* 16, 6849–6859
  33. Baranova, S., Tuzikov, F. V., Zakharova, O. D., Tuzikova, N. A., Calmels, C., Litvak, S., Tarrago-Litvak, L., Parissi, V., and Nevinsky, G. A. (2007) *Nucleic Acids Res.* 35, 975–987
  34. Johnson, A. A., Santos, W., Pais, G. C., Marchand, C., Amin, R., Burke, T. R., Jr., Verdine, G., and Pommier, Y. (2006) *J. Biol. Chem.* 281, 461–467
  35. Dicker, I. B., Samanta, H. K., Li, Z., Hong, Y., Tian, Y., Banville, J., Remillard, R. R., Walker, M. A., Langley, D. R., and Krystal, M. (2007) *J. Biol. Chem.* 282, 31186–31196
  36. Zhao, Z., McKee, C. J., Kessl, J. J., Santos, W. L., Daigle, J. E., Engelman, A., Verdine, G., and Kvaratskhelia, M. (2008) *J. Biol. Chem.* 283, 5632–5641
  37. Lapadat-Tapolsky, M., de Rocquigny, H., Van Gent, D., Roques, B., Plasterk, R., and Darlix, J. L. (1993) *Nucleic Acids Res.* 21, 831–839
  38. Carreau, S., Batson, S. C., Poljak, L., Mouscadet, J. F., de Rocquigny, H., Darlix, J. L., Roques, B. P., Kas, E., and Auclair, C. (1997) *J. Virol.* 71, 6225–6229
  39. Delelis, O., Lehmann-Che, J., and Saib, A. (2004) *Curr. Opin. Microbiol.* 7, 400–406
  40. Yu, S. F., Baldwin, D. N., Gwynn, S. R., Yendapalli, S., and Linial, M. L. (1996) *Science* 271, 1579–1582
  41. Li, M., Mizuuchi, M., Burke, T. R., Jr., and Craigie, R. (2006) *EMBO J.* 25, 1295–1304
  42. Ren, G., Gao, K., Bushman, F. D., and Yeager, M. (2007) *J. Mol. Biol.* 366, 286–294
  43. Bosserman, M. A., O'Quinn, D. F., and Wong, I. (2007) *Biochemistry* 46, 11231–11239
  44. Zhou, X. Z. (1995) *Biophys. J.* 69, 2298–2303
  45. Harding, S. E., Horton, J. C., Jones, S., Thornton, J. M., and Winzor, D. J. (1999) *Biophys. J.* 76, 2432–2438



## II.C. Conclusion

Nous avons mis au point une technique de marquage enzymatique spécifique de l'IN de PFV-1 au niveau d'une étiquette rajoutée en C-ter de la séquence primaire protéique à l'aide d'un fluorophore de petite taille (le TAMRA). L'étude de l'IN de PFV-1, marquée au TAMRA ou non, par FCS et TFA montrent une même transition monomère – dimère avec une concentration de demi-transition de l'ordre de 20 à 30  $\mu\text{M}$ . Ces résultats montrent d'une part que le marquage ne modifie pas la solubilité et l'état multimérique de l'IN de PFV-1 seule en solution et d'autre part que la solubilité de l'IN de PFV-1 est plus importante que celle de l'IN du VIH-1. Dans des conditions équivalentes, l'IN du VIH-1 est sous forme tétramérique pour des concentrations en IN inférieures à 200 nM et dans un état agrégé pour des concentrations supérieures à 200 nM ((Leh *et al.*, 2000); (Laboulais *et al.*, 2001)).

Des mesures de FRET entre l'IN de PFV-1 marquée et des ADN de différentes longueurs couplés à la fluorescéine au niveau de leurs extrémités 5' ou 3' ont révélées une tendance de l'IN à reconnaître préférentiellement l'extrémité qui va être clivée lors du 3'-processing. Cette préférence de l'IN de PFV-1 pour l'extrémité processée de l'ADN viral a été confirmée par des expériences de compétitions entre des ADN mimant cette extrémité et des ADN non spécifiques. Cette observation n'avait jusque là jamais pu être mise en évidence pour l'IN du VIH-1 pour laquelle seule l'activité catalytique révélait la nature des complexes formés ((Esposito *et al.*, 1998); (Guiot *et al.*, 2006)). Les expériences de FRET montrent également que l'IN de PFV-1 établit des interactions non spécifiques au niveau de séquences positionnés en interne de l'ADN. Cette capacité de liaison non spécifique à l'ADN pourrait être dû, comme c'est le cas pour l'IN du VIH-1, à un mode d'interaction non spécifique avec l'ADN, par exemple *via* le site qui sert à fixer l'ADN cible.

L'IN de PFV-1 nous a également permis de mettre en évidence deux types de complexes peu ou pas actifs de poids élevé selon la taille de l'oligonucléotide clivé. Sur les ADN courts ( $\leq 21$  paires de bases), une sélection des complexes catalytiquement compétents est possible par une augmentation de la force ionique. Ce paramètre permet de moduler les interactions non spécifique IN / IN qui aboutissent à la formation de complexes de haut poids moléculaires peu actifs, identiques à ceux observés précédemment avec l'IN du VIH-1 (Guiot *et al.*, 2006). La caractérisation par TFA des complexes catalytiquement compétents montre

que la forme multimérique active de l'IN de PFV-1 pour le 3'-processing est, comme pour l'IN du VIH-1, un dimère ((Faure *et al.*, 2005); (Guiot *et al.*, 2006)). Sur des ADN plus long ( $\geq 45$  paires de bases), l'augmentation de la force ionique ne permet pas de sélectionner les complexes catalytiquement compétents. Les interactions mises en jeu sont donc de nature différente. Ces complexes de haut poids moléculaires sont associés par l'intermédiaire d'interactions IN / ADN plus stables que les interactions non spécifiques IN / IN. La longueur importante de l'ADN permet la fixation de plusieurs IN à la suite les uns des autres telles les perles d'un collier. Cette interaction de l'IN de PFV-1 avec des séquences internes de l'ADN est confirmée par les expériences de FRET. *In vivo*, la fixation de l'IN au niveau de séquences internes de l'ADN viral doit être limitée par la présence de la nucléocapside qui recouvre l'ADN en totalité avec une forte affinité à l'exception des extrémités (Lapadat-Tapolsky *et al.*, 1993). La NC pourrait donc permettre de manière indirecte le positionnement correct de l'IN au niveau des extrémités processables de l'ADN viral. Dans le cas de PFV-1, aucune NC ne contenant de doigt de zinc caractéristique n'a été mise en évidence. Cependant, au niveau de la région C-terminale de Gag se trouvent trois motifs appelés boîtes « GR » (motif Gly-Arg) qui sont supposés avoir le même rôle que la NC du VIH-1 (Delelis *et al.*, 2004).

Le test d'activité basé sur l'anisotropie statique de fluorescence nous a permis également de déterminer d'une part, les affinités de ces deux IN pour leur ADN viral respectif (130 nM pour l'IN de PFV-1 et 40 nM pour l'IN du VIH-1 (Guiot *et al.*, 2006)) et d'autre part, leur constante catalytique de 3'-processing respective ( $0.0028 \text{ min}^{-1}$  pour l'IN de PFV-1 et  $0.004 \text{ min}^{-1}$  pour l'IN du VIH-1 (Guiot *et al.*, 2006)). Cependant, malgré ces différences, les propriétés enzymatiques de ces deux IN rétroviraux sont proches puisqu'elles catalysent toutes les deux la même réaction séquentielle de coupure endonucléolytique spécifique et de transfert de brin. Ceci confirme le modèle précédemment établi pour l'IN du VIH-1 ((Guiot *et al.*, 2006); (Smolov *et al.*, 2006)) : la cinétique lente « single turn-over » observée pour le 3'-processing n'est pas seulement due à un état agrégé de l'IN, mais à un processus lent mettant en jeu une probable transition conformationnelle de l'IN sur l'ADN viral nécessaire pour l'activation du complexe IN / ADN.

### ***III. Etude différentielle selon le contexte cationique au site actif de l'implication du domaine N-terminal de l'IN du VIH-1 lors du processus multimérique de fixation à l'ADN viral***

#### **III.A. Introduction**

L'émergence de virus résistant lors des traitements anti-VIH de type HAART, dues à la plasticité de son génome, montre que les efforts menés actuellement pour découvrir de nouveaux anti-rétroviraux doivent être poursuivis. L'IN représente une cible attractive d'une part parce que son activité est nécessaire au bon déroulement du cycle de réplication viral et d'autre part parce qu'il n'existe aucune enzyme équivalente au niveau cellulaire. De plus, elle permettrait d'agir sur une cible encore non utilisée par la pharmacopée actuelle qui cible principalement les deux autres enzymes rétrovirales que sont la RT et la P. La diversification des cibles d'action permet de limiter le phénomène de résistance croisée du virus face aux différentes molécules administrées. Depuis de nombreuses années, plusieurs tests d'activité ont été mis en place *in vitro* afin de trouver des inhibiteurs de l'IN et de comprendre leur mécanisme d'action. Il faut également noter que les tests d'activité menés en présence de composés anti-IN et de  $Mg^{2+}$  comme cofacteur cationique sont généralement plus prédictifs, qu'en présence de  $Mn^{2+}$ , des activités anti-IN trouvées ensuite lors de tests menés *ex vivo*. A ce jour, les molécules inhibitrices de l'IN se divisent en deux classes : les inhibiteurs compétitifs qui interagissent au niveau du site de fixation de l'ADN cible. Ces composés inhibent le transfert de brin et dans une moindre mesure le 3'-processing. Ce premier groupe est qualifié d'inhibiteur du transfert de brin de l'IN (INSTI). L'un d'eux, le Raltegravir a été approuvé par la FDA en 2007. La deuxième classe comprend des composés qui sont des inhibiteurs compétiteurs de l'ADN viral. Lorsqu'ils sont fixés sur le site donneur, ils inhibent le 3'-processing. Cette deuxième famille appelée inhibiteur de la fixation à l'ADN (INBI) comprend notamment des molécules de la famille des styrylquinolines. Pour chacune de ces deux familles, des mutations de résistances au niveau de la séquence de l'IN ont déjà été identifiées.

Etant donné que l'IN possède une activité catalytique dépendante de son état multimérique : *vivo* pour le 3'-processing et au minimum tétramérique pour l'intégration concertée, des inhibiteurs modifiant les équilibres multimériques de l'IN pourraient former une nouvelle classe d'inhibiteur. Ces molécules agissant comme des inhibiteurs de type

allostériques, ils ne devraient présenter aucune résistance croisée avec les InBI et les InSTI. Depuis peu, des peptides qui perturbent l'état oligomérique de l'IN ont été identifiés et sont appelés « shiftides » de par leur capacité à favoriser la formation de tétramère au détriment du monomère et du dimère compétent pour le 3'-processing (Hayouka *et al.*, 2007). Cette approche prometteuse requiert cependant une bonne compréhension des rôles respectifs des différents domaines de l'IN dans ce processus de multimérisation.

Les déterminants de la multimérisation de l'IN du VIH-1 ne sont pas encore clairement identifiés : le CC, qui contient la triade catalytique et interagit avec l'ADN, montre dans toutes ses structures cristallographiques (IN<sup>1-212</sup>, IN<sup>55-288</sup> et IN<sup>55-212</sup>) une surface d'interaction entre les deux monomères formant le dimère. Cette surface de par son importance et sa constance est sûrement biologiquement relevante. En ce qui concerne les domaines N-ter et C-ter, différentes études montrent leur implication dans le processus de multimérisation de l'IN. Cependant, leur implication principale pourrait se faire dans deux cadres d'organisation multimérique différents. La structure du double domaine IN<sup>1-212</sup> montre une surface d'interaction entre les deux monomères contenus dans la maille cristallographique composée de deux zones, une correspondant à l'interaction entre les deux CC et une autre correspondant à la surface d'interaction entre les deux domaines N-ter. Cette interface n'est pas totalement identique avec celle observée pour la structure du domaine N-ter seul. Le domaine N-ter pourrait donc stabiliser et favoriser la formation du dimère d'IN. De plus, ce domaine ne peut interagir avec l'ADN qu'en présence du CC. Il faut noter également que la structuration du domaine N-ter, qui contient un pseudo-doigt de zinc, en présence de Zn<sup>2+</sup> permet d'une part de promouvoir la multimérisation de l'IN *in vitro* et d'autre part de stimuler l'activité de 3'-processing en présence de Mg<sup>2+</sup>, qui nécessite une architecture dimérique de l'IN. De plus, les mutations du pseudo-doigt de zinc aboutissent à des virus non infectieux incapables d'intégrer le génome viral au sein de la cellule hôte. Le domaine N-ter pourrait donc jouer un rôle prépondérant dans l'organisation architecturale du dimère d'IN responsable de l'activité de 3'-processing. Le domaine C-ter qui, quant à lui, présente un repliement de type SH3, généralement impliqué dans des interactions protéines / protéines, forme des contacts dits « non spécifiques » avec l'ADN. Au niveau de la structure cristallographique du double domaine CC / C-ter et contrairement à la structure du domaine C-ter seul, les deux domaines C-ter n'interagissent pas entre eux au niveau du dimère constituant de la maille cristallographique, par contre ils établissent des interactions avec les domaines C-ter d'autres dimères. Il est intéressant de noter que la mutation du résidu W235 du domaine C-ter, est

silencieuse lors des tests d'activité de 3'-processing et de transfert de brin de l'IN. Cependant, l'effet sur l'intégration concertée est différent selon le résidu substituant, si l'hydrophobicité est conservée (W235F), la mutation reste silencieuse, le rendement de l'intégration concertée est même amélioré. De plus, cette mutation n'influe pas l'infectiosité des virus (Engelman *et al.*, 1995). Par contre, la mutation W235A abolit d'une part l'intégration concertée (Li *et al.*, 2005) et d'autre part, l'infectiosité des virions (Cannon *et al.*, 1994). Ces différentes considérations montrent que le domaine C-ter serait plutôt impliqué *in vivo* des dimères d'IN. Afin de déterminer le rôle du domaine N-ter dans le processus de multimérisation dans le cadre du 3'-processing, nous avons utilisé une molécule connue comme étant un éjecteur du zinc des doigts de zinc : le DIBA-1 (Figure 1). Ce composé agit notamment sur la nucléocapside du VIH-1 et sur un récepteur aux œstrogènes.

### III.B. Matériel et méthodes

#### Oligonucléotides.

Les oligonucléotides **U5A** (5'-GTG TGG AAA ATC TCT AGC AGT) et son brin complémentaire **U5B** (5'-ACT GCT AGA GAT TTT CCA CAC), mimant l'extrémité U5 de l'ADN viral du VIH-1, sont utilisés pour le test d'activité de 3'-processing. Pour les expériences de fixation à l'ADN, U5A est marquée à la fluorescéine au niveau de l'extrémité 3', il est alors nommé U5A-3'-F. Le test d'activité de désintégration est réalisé sur un substrat possédant une conformation en Y nommé **D** (5'-TGC TAG TTC TAG CAG GCC CTT GGG CCG GCG CTT GCG CC) déjà décrit ((Leh *et al.*, 2000); (Chow, 1997)). Les constructions des IN tronquées d'un ou deux domaines ( $\Delta$ N,  $\Delta$ C et CC) sont générées par PCR en utilisant les oligonucléotides suivants :  **$\Delta$ NPCR** 5'-CAT ATG CAT GGA CAA GTA GAC TG; **T7Term** 5'-TGC TAG TTA TTG CTC AGC GG;  **$\Delta$ CPCR** 5'-CTA TAA TTC TTT TAG TTT GTA; **T7Prom** 5'-TAA TAC GAC TCA CTA TAG G. Les oligonucléotides marqués ou non avec la fluorescéine sont commandés chez Eurogentec (Liege, Belgique) et purifiés dès leur réception par électrophorèse sur un gel dénaturant acrylamide / urée 15%.

#### Production et purification des intégrases entière et tronquées.

Toutes les protéines (excepté IN<sup>1-213</sup>) contiennent la mutation C280S de manière à diminuer les ponts disulfures intermoléculaires, car il a été montré précédemment que le résidu C280 est responsable de la présence de multimères covalents lorsque l'on est dans des conditions sans agent réducteurs (Bischerour *et al.*, 2003). Les mutations C56S, C65S et C56S-C65S ont été introduites, dans le contexte du mutant C280S, dans le plasmide pET-15b-IN (Leh *et al.*, 2000) codant pour l'IN du VIH-1 avec au niveau de son extrémité N-ter une étiquette 6 Histidines. Les mutagenèses dirigées

furent réalisées avec le kit de Stratagene : Quickchange® II Site-directed Mutagenesis et vérifiées par séquençage (Genome express). Les constructions tronquées d'un ou de deux domaines, IN<sup>50-288</sup> (ou ΔN), IN<sup>1-213</sup> (ou ΔC) et IN<sup>50-213</sup> (ou CC) furent générées par PCR à partir de l'IN entière portant la mutation C280S. Les amorces (ΔNPCR et T7Term pour ΔN; T7Prom et ΔCPCR pour ΔC; ΔNPCR et ΔCPCR pour CC) furent déterminées de manière à générer des fragments d'ADN contenant un site de restriction spécifique de NdeI à l'extrémité 5' et un codon stop flanqué par un site spécifique de BamHI à l'extrémité 3'. Les fragments produits par la PCR furent dans un premier temps introduits dans le plasmide pGEM®-T easy vector (Promega). Les séquences codant pour les différentes protéines tronquées furent ensuite clonées dans le plasmide pET-15b après une double digestion par les enzymes de restriction NdeI et BamHI.

Les différentes protéines portant l'étiquette 6His furent surexprimées dans des bactéries *E. coli* BL21-codonplus® (DE3+) (Stratagene®) puis purifiées selon le protocole précédemment décrit (Leh et al., 2000) avec plusieurs modifications, que nous allons exposer rapidement ci-dessous. Les cultures bactériennes furent incubées à 37°C et l'expression des protéines recombinantes induites par ajout d'IPTG (à la concentration finale de 0.25 mM) lorsqu'elles ont atteint une DO<sub>600nm</sub> ≈ 0,7. La température est alors abaissée à 25°C et les cultures sont incubées sous agitation pendant 5 heures. La purification est réalisée en utilisant des billes NiNTA (Qiagen). L'agent réducteur (β-mercaptoéthanol) est éliminé durant les étapes de lavages et d'élutions ainsi que pendant les dialyses suivantes. Tous les tampons de purification furent additionnés de 50 μM final de ZnSO<sub>4</sub> sauf ceux des protéines que nous avons utilisées pour les expériences d'éjection du Zn<sup>2+</sup>. Les protéines sont ensuite aliquotées, puis rapidement congelées dans de l'azote liquide et conservées à -80°C.

### **Réactions de 3'-processing et de désintégration. Détermination des valeurs d'IC<sub>50</sub>.**

Pour les tests d'activité, 100 pmol de chaque oligonucléotide U5A (substrat du 3'-processing) ou D (substrat de la désintégration) sont marqués radioactivement par la T4 polynucléotide kinase et 50 μCi d'[γ-<sup>32</sup>P]ATP (3000 Ci/mmol). La kinase T4 est ensuite inactivée par la chaleur et les nucléotides non incorporés sont éliminés par filtration sur une colonne Sephadex G-25 (GE Healthcare). Pour la réaction de 3'-processing, on ajoute au mélange, de manière équimolaire à U5A, le brin complémentaire U5B et une concentration finale de 100 mM de NaCl. Le mélange est incubé à 85°C pendant 5 min puis l'hybridation spécifique s'effectue lors du refroidissement lent du mélange à température ambiante. Les réactions de 3'-processing ou de désintégration sont réalisées en utilisant respectivement 1 nM du duplex U5A/U5B ou d'ADN D dans un tampon contenant 20 mM Tris pH 7.0, 15% DMSO (v/v), 50 mM NaCl, 10 mM de cation divalent (Mg<sup>2+</sup> ou Mn<sup>2+</sup>), l'IN entière ou les protéines tronquées (les concentrations sont indiquées dans la légende de chaque figure) et différentes concentrations de DIBA-1. Les mélanges réactionnels sont incubés pendant 3 heures à 37°C, puis la réaction est arrêtée par ajout de 80 μL d'un tampon contenant 6 mM d'EDTA, 0.125 mg/ml de

glycogène, 400 mM de NaOAc, on réalise ensuite une extraction au phénol-chloroforme. Les fragments d'ADN sont ensuite précipités à l'éthanol, puis resuspendus dans un tampon de charge contenant 20 mM d'EDTA, 80% de formamide, 0.05% de bleu de bromophénol et 0.05% de xylène cyanol. Ils sont ensuite déposés sur un gel d'électrophorèse acrylamide / urée dénaturant de 12%. Les gels sont analysés avec le STORM 840<sup>TM</sup> PhosphorImager (Molecular Dynamics, Sunnyvale, CA, USA) et quantifiés en utilisant le logiciel Image Quant<sup>TM</sup> 4.1. La valeur IC<sub>50</sub> de la courbe d'inhibition représente la concentration de drogue donnant 50% d'inhibition (ou l'activité relative = 50%).

### **Test de fixation de l'IN sur l'ADN.**

L'interaction entre l'IN et l'oligonucléotide de 21-mer double brin marqué à la fluorescéine est détecté par anisotropie de fluorescence statique avec le Beacon 2000 (PanVera, Madison, USA). L'oligonucléotide double brin est obtenu en mélangeant de manière équimolaire les deux brins complémentaires U5B et U5A-3'-F dans un tampon Tris-HCl à 20 mM (pH 7.2) contenant 100 mM de NaCl. Le mélange est chauffé à 85°C pendant 5 min puis l'hybridation spécifique s'effectue lors du refroidissement lent du mélange à température ambiante. Pour déterminer la valeur du K<sub>d</sub> (K<sub>d,app</sub>), l'ADN marqué à la fluorescéine (1 nM) est incubé avec des concentrations croissantes d'IN (entière ou protéines tronquées) dans un tampon Tris 20 mM (pH 7.2) contenant 15% DMSO (v/v), 50 mM de NaCl et 10 mM de MnCl<sub>2</sub> ou MgCl<sub>2</sub> pendant 20 min à 25°C, et l'anisotropie statique (r) est ensuite mesurée. En présence de DIBA-1, les expériences de titrations sont réalisées avec un ratio constant DIBA : IN (from 1:1 to 3:1). La fraction de saturation Y est calculée avec la formule suivante:  $Y = (r - r_{\text{libre}}) / (r_{\text{lié}} - r_{\text{libre}})$ , où r<sub>lié</sub> et r<sub>libre</sub> représentent respectivement l'anisotropie de l'ADN lié et libre (aucun changement significatif d'intensité n'est observé). Le nombre de Hill, ñ, est calculé en fittant directement la courbe de titration avec la fonction de Hill du logiciel origin 6.0. K<sub>d,app</sub> ou IN<sub>50</sub> représente la concentration d'IN requise pour titrer l'ADN à demi-saturation.

### **Détermination de l'éjection du zinc par le DIBA-1.**

L'éjection du zinc de l'IN est quantifiée en utilisant un agent chélateur coloré : le 4-(2-pyridylazo)resorcinol (PAR) (Sigma). La libération du Zn<sup>2+</sup> par action du DIBA-1 est suivie par absorbance avec un spectrophotomètre Uvikon 941 (Kontron) car la formation du complexe Zn<sup>2+</sup>-(PAR)<sub>2</sub> induit une augmentation de l'absorbance à 495 nm. Toutes les expériences ont été réalisées dans un tampon Tris (20 mM, pH 7.0) contenant 20% DMSO (v/v) de manière à augmenter la solubilité du DIBA-1. Le coefficient d'extinction molaire du complexe Zn<sup>2+</sup>-(PAR)<sub>2</sub> fut mesuré en présence de DMSO (20 %) en utilisant des solutions de zinc de concentration connue (de 1 à 6 µM). Nous avons trouvé un coefficient d'extinction molaire de 6.1 10<sup>4</sup> M<sup>-1</sup> cm<sup>-1</sup> dans 20% DMSO, cette valeur est légèrement plus faible que celle trouvée dans l'eau précédemment (6.6 10<sup>4</sup> M<sup>-1</sup> cm<sup>-1</sup>) (Hunt et al., 1984).

### III.C. Résultats

#### **Inhibition différentielle par le DIBA des activités de l'IN $Mg^{2+}$ - et $Mn^{2+}$ -dépendantes.**

Les activités de 3'-processing et de désintégration catalysée par l'IN entière sont inhibées de manière efficace par le DIBA-1 avec des valeurs d'IC50 de 11-13 nM pour les deux réactions en présence de  $Mg^{2+}$  (Fig 1, partie A-D). Il est intéressant de noter que ces valeurs sont significativement plus élevées en présence de  $Mn^{2+}$ , *i.e.* 118-123 nM. Toutes les réactions sont réalisées sans agent réducteur afin d'éviter une réduction du pont disulfure du DIBA-1 (Fig 2). D'ailleurs, l'addition de DTT ou de  $\beta$ -mercaptoéthanol dans le mélange réactionnel abolit totalement l'effet inhibiteur du DIBA-1 (Fig 1, partie E). Nous avons ensuite réalisé des tests de fixation de l'IN sur l'ADN afin de savoir si le DIBA agit plutôt comme un inhibiteur compétitif de l'ADN viral (InBI) en inhibant les interactions IN / ADN comme par exemple les composés de la famille des styrylquinolines (Deprez *et al.*, 2004), ou comme un inhibiteur de l'étape catalytique qui suit l'étape de fixation à l'ADN.

#### **Le DIBA-1 n'inhibe pas les interaction IN / ADN.**

L'hypothèse selon laquelle le DIBA peut inhiber l'activité de l'IN en inhibant les interactions IN / ADN fut vérifiée à l'aide de l'anisotropie de fluorescence statique (Guiot *et al.*, 2006). Des concentrations croissantes d'IN entière ont été ajoutées sur un oligonucléotide marqué à la fluorescéine qui mime l'extrémité U5 de l'ADN du VIH-1 en absence (Fig 3, carrés noirs) ou en présence (carrés blancs) de DIBA-1. Les mélanges réactionnels réalisés en présence de DIBA contiennent un ratio constant DIBA : IN (2 : 1). Il apparaît clairement que la fixation de l'IN sur l'ADN n'est pas affectée en présence de DIBA-1. Pour chacune des conditions, l'activité de 3'-processing est reportée sur le graphe et confirme que le DIBA inhibe fortement cette activité (Fig 3, comparaison des cercles blancs et noirs) dans des conditions pour lesquelles il n'exerce aucune influence sur la formation des complexes IN / ADN. De ce fait, en présence de DIBA-1, l'IN interagit avec l'ADN mais forme des complexes qui ne sont catalytiquement pas compétents, et ce quelque soit le cation divalent présent au niveau du site actif :  $Mg^{2+}$  (Fig 3) ou  $Mn^{2+}$  (non montré) (voir Tableau 2 pour l'ensemble des valeurs des affinités de l'IN pour l'ADN en présence des deux cations).

#### **Influence du DIBA-1 sur la fixation coopérative de l'IN sur son substrat ADN.**

Il a été montré précédemment que la fixation de l'ADN en présence de  $Mg^{2+}$  est coopérative (Deprez *et al.*, 2004). Ce résultat est confirmé dans cette étude : la fixation de l'IN sur l'ADN est caractérisée par un nombre de Hill ( $\tilde{n}$ ) de 2 (Tableau 1). Le DIBA-1 affecte de manière significative le nombre de Hill qui passe de 2 à 1.5 (Fig 4 et Table 1). Ceci suggère que le DIBA-1 interfère avec la fixation coopérative de l'IN sur l'ADN. Il est important de noter que la fixation de l'IN sur l'ADN est moins coopérative en présence de  $Mn^{2+}$  ( $\tilde{n} = 1.29$ ) et que le DIBA-1 n'affecte pas de manière significative cette valeur ( $\tilde{n} = 1.20$ ) (Tableau 1). Les affinités apparentes ( $K_{d,app}$ ) pour l'IN entière sont similaires quelque soit le cofacteur cationique présent et ce indépendamment de la présence de DIBA-1. Ces données confirment que le DIBA-1 n'influe pas sur la formation des complexes IN / ADN (Tableau 2, première ligne). Par contre, dans le contexte  $Mg^{2+}$ , il semble que la nature des interactions IN / IN sur l'ADN substrat ou le statut oligomérique des complexes IN / ADN sont différents selon la présence ou l'absence de DIBA-1.

### **Le DIBA-1 induit une éjection de zinc de l'IN.**

Sachant que le zinc est décrit dans la littérature comme étant un modulateur de l'état oligomérique de l'IN ((Zheng *et al.*, 1996); (Lee *et al.*, 1997)), l'éjection du zinc induite par le DIBA-1 a été testée et mesurée par spectrophotométrie. Pour cela nous avons utilisé un agent chélateur du zinc coloré, le 4-(2-pyridylazo)resorcinol (PAR). Comme cela est montré sur la Fig 4A, le DIBA-1 induit une libération de zinc dans le milieu réactionnel. Cet effet est annihilé par l'ajout d'un agent réducteur comme le  $\beta$ -mercaptoéthanol. D'après le coefficient d'extinction molaire du complexe  $Zn^{2+}$ -(PAR)<sub>2</sub> dans 20% DMSO ( $6.1 \cdot 10^4 \text{ M}^{-1}\text{cm}^{-1}$ ) et la concentration en IN (1  $\mu\text{M}$ ), l'absorbance trouvée au plateau (0.055) indique que l'IN contient du zinc à hauteur de 0.9  $Zn^{2+}$  par monomère (Fig 5A). Cette valeur est en accord avec celle attendue pour une protéine contenant un doigt de zinc de type HHCC. Nous avons, de plus, trouvé une équivalence de deux molécules de DIBA par monomères d'IN pour éjecter l'ion  $Zn^{2+}$  du pseudo-doigt de zinc (Fig 5B).

### **Le DIBA-1 inhibe l'activité de désintégration des IN tronquées.**

Afin de déterminer les domaines de l'IN impliqués dans la réponse inhibitrice du DIBA-1, nous avons testé son effet sur plusieurs IN tronquées que sont les deux IN doubles domaines  $IN^{1-213}$  ( $\Delta C$ ) et  $IN^{50-288}$  ( $\Delta N$ ) et le core catalytique  $IN^{50-213}$  (CC). Une étude comparative sur l'activité et l'effet différentiel du DIBA-1 sur ces trois protéines tronquées et

sur l'IN entière a été menée en utilisant un test de désintégration car les trois protéines tronquées ne sont pas compétentes pour la catalyse de la réaction de 3'-processing, et ce quelque soit le cofacteur cationique. Il faut également noter que ces protéines tronquées ne sont compétente pour l'activité de désintégration qu'en présence de  $Mn^{2+}$ . Comme cela est montré sur la Fig 6, toutes les protéines tronquées sont sensibles au DIBA-1, cependant, leurs  $IC_{50}$  sont plus élevés que celui obtenu pour l'IN entière. Ce résultat met en lumière un autre effet du DIBA-1, *i.e.* le DIBA-1 agit sur le CC par un mécanisme indépendant de l'éjection du zinc.

De manière à comprendre pourquoi les protéines tronquées sont incompetentes pour le 3'-processing, nous avons comparé l'affinité des protéines tronquées ( $\Delta N$ ,  $\Delta C$  et CC) pour l'ADN avec celle de l'IN entière (Tableau 2). Toutes les protéines testées présentent une capacité de fixation à l'ADN, ce qui suggère qu'elles sont seulement déficientes pour l'étape catalytique. Sur le Tableau 2, il apparaît que le domaine C-ter joue un rôle clef dans l'affinité du complexe IN / ADN, et ce seulement dans un contexte  $Mg^{2+}$ . En effet, les valeurs de  $K_{d,app}$ , obtenus en présence de  $Mn^{2+}$ , sont similaires pour les quatre protéines. A l'inverse, en présence de  $Mg^{2+}$ , les IN délétées du domaine C-ter présentent une affinité fortement diminuée pour l'ADN. Ceci confirme l'importance du rôle joué par le domaine C-ter dans la stabilisation des complexes IN / ADN. Ces données suggèrent également que ce domaine pourrait former des contacts spécifiques avec l'ADN puisque le contexte  $Mg^{2+}$  est plus propice à révéler les contacts de nature spécifique.

### **Le site catalytique est également ciblé par le DIBA-1.**

Les résultats obtenus sur les protéines tronquées montrent que le DIBA-1 exerce un effet inhibiteur qui n'est pas strictement restreint à la présence du domaine N-ter contenant le doigt de zinc. C'est pour cette raison que l'activité de désintégration du CC est également inhibée par le DIBA-1. Trois résidus cystéines (Cys 56, 65 et 130) du CC sont des cibles potentielles d'interaction pour le DIBA-1. La Cys130 étant critique pour la stabilité de l'IN entière, sa mutation n'a pas été réalisée. Les mutants C56S, C65S et C56S-C65S ont été réalisés à partir d'une construction de l'IN contenant déjà la mutation C280S. Leur sensibilité respective au DIBA-1 a été testée en présence des deux cofacteurs cationiques. Tout d'abord en présence de  $Mn^{2+}$ , le mutant C65S/C280S présente une  $IC_{50}$  comparable à celle de l'IN C280S qui sert de témoin (Tableau 3). Par contre les valeurs d' $IC_{50}$  augmentent de manière significative pour les mutants C56S/C280S et C56S/C65S/C280S ce qui suggère que ces

protéines sont moins affectées par le DIBA-1 et donc que le résidu C56 est la cible principale du DIBA-1 dans ce contexte. Ce résultat permet d'expliquer la sensibilité du CC au DIBA-1 et montre donc que l'éjection du zinc, présent au niveau du domaine N-ter, n'est pas le seul effet induit par cette molécule.

Il est particulièrement intéressant de noter que les résultats obtenus en présence de  $Mg^{2+}$  sont différents. En effet, tous les mutants sont inhibés avec des efficacités comparables et les  $IC_{50}$  calculées sont meilleures qu'en présence de  $Mn^{2+}$  (Tableau 3). Ceci suggère qu'en présence de  $Mg^{2+}$ , l'effet du DIBA-1 sur le domaine N-ter est dominant sur celui exercé sur le résidu C56. Cette hypothèse permet d'expliquer la différence d' $IC_{50}$  pour l'IN entière entre les deux contextes cationiques. De plus, contrairement à ce que nous avons vu précédemment en présence de  $Mg^{2+}$ , les effets du DIBA-1 qui induisent l'éjection du zinc et donc de manière indirecte la coopérativité de la fixation de l'IN sur l'ADN, ne semblent pas jouer un rôle crucial en présence de  $Mn^{2+}$ . Les performances catalytiques deviennent alors plus sensibles aux perturbations induites au niveau du site actif. L'activité  $Mn^{2+}$  dépendante est donc principalement affectée par un effet direct du DIBA-1 au niveau du site actif.

### III.D. Discussion

Nous avons montré que le DIBA-1 inhibe de manière efficace les activités de 3'-processing et de désintégration de l'IN du VIH-1 avec des  $IC_{50}$  de l'ordre de 15 nM en  $Mg^{2+}$  et 100 nM en  $Mn^{2+}$ . Cet effet anti-IN est clairement distinct de ceux observés pour les deux principales familles d'inhibiteurs de l'IN. Les InBI, qui représentent actuellement la famille qui inhibent plus efficacement le 3'-processing comparée au transfert de brin, sont des compétiteurs de la fixation de l'IN sur l'ADN viral. Au contraire, les InSTI ne peuvent pas interagir au niveau du site actif de l'IN libre et n'inhibent pas la réaction de 3'-processing. Ce sont des molécules qui ciblent de manière spécifique les complexes IN / ADN viral déjà établis, et qui inhibent la fixation de l'ADN cible et donc la réaction de transfert de brin. Les titrations réalisées en présence ou non de DIBA-1 montrent clairement que ce composé ne modifie pas l'affinité globale de l'IN pour l'ADN. Le mécanisme d'inhibition de l'IN par le DIBA-1 est donc de type non compétitif de l'ADN viral. Nos résultats suggèrent la présence de deux modes d'inhibition de l'IN : un dépend de l'éjection du zinc coordonné au niveau du pseudo-doigt de zinc du domaine N-ter, l'autre est indépendant et concerne directement le site actif du CC. Nos résultats suggèrent que la prévalence de l'un ou de l'autre mécanisme dépend du cofacteur cationique lié à l'IN.

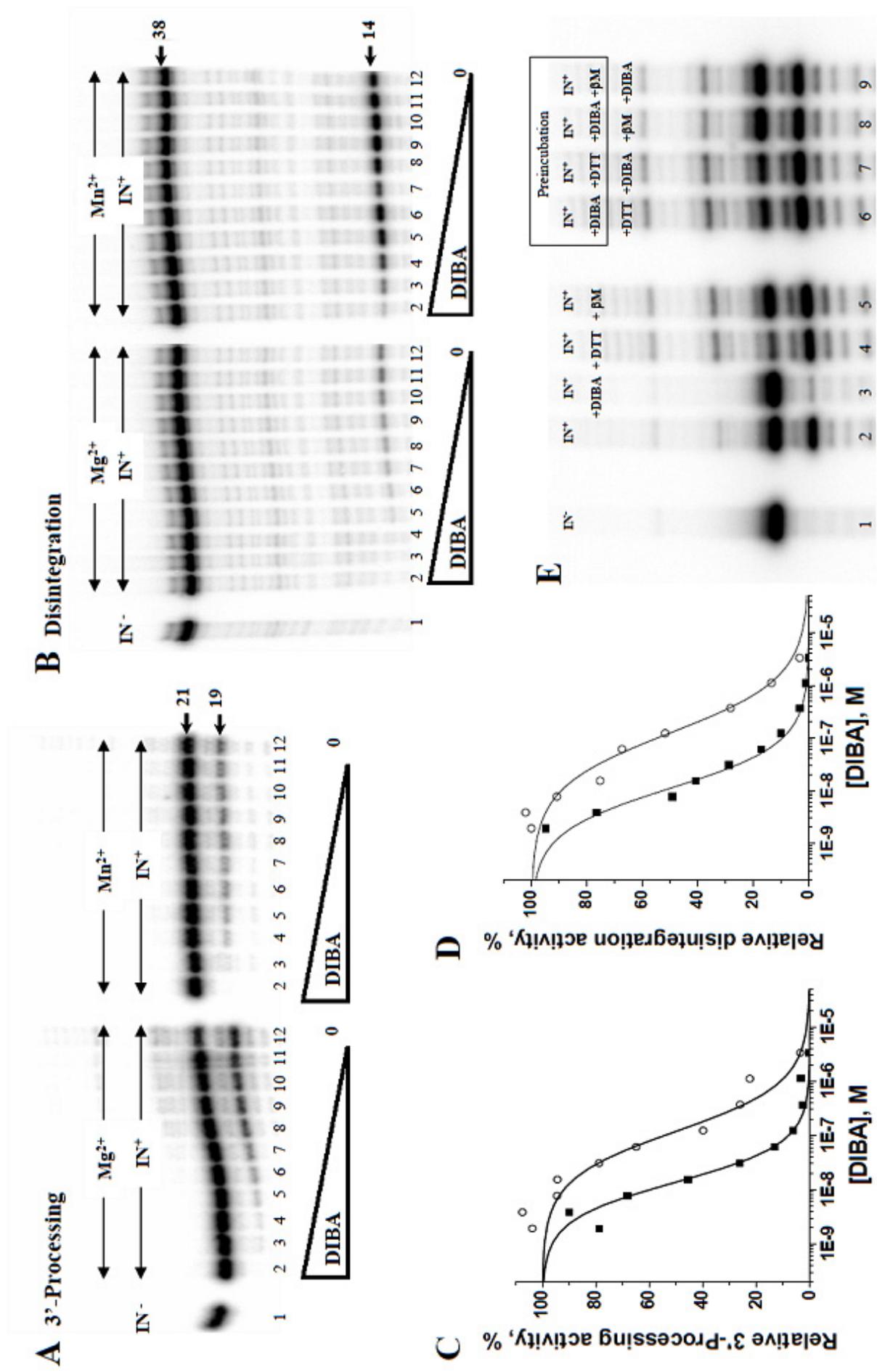
Les données préliminaires de spectrométrie de masse nous indiquent la formation d'une interaction stable entre le DIBA-1 et la cystéine 40 du doigt de zinc (collaboration avec H. Leh, LBPA). Cette interaction s'accompagne de l'éjection du  $Zn^{2+}$  dans le milieu réactionnel. De plus, les titrations réalisées en présence de différents ratio DIBA : IN montrent que le DIBA-1 induit une perte de la coopérativité de la fixation de l'IN sur l'ADN viral en présence de  $Mg^{2+}$ . En présence de  $Mn^{2+}$ , la coopérativité est de manière significative plus faible et demeure insensible à la présence de DIBA-1. En tenant compte du fait que (i) la multimérisation est cruciale pour l'activité de l'IN, (ii) le zinc coordonné par le doigt de zinc du domaine N-ter de l'IN est décrit comme jouant un rôle clef dans le processus de multimérisation de l'IN et (iii) le DIBA éjecte de manière efficace le zinc de l'IN par une interaction avec la C40 qui fait partie des quatre résidus impliqués dans la coordination du zinc, nous pouvons raisonnablement attribuer à l'éjection du zinc l'inhibition de l'activité de l'IN. L'ensemble de ces données montre également que la réaction de 3'-processing nécessite une dimérisation coopérative de l'IN dans un contexte  $Mg^{2+}$ . Ce qui n'est pas valable dans le contexte  $Mn^{2+}$ .

Les expériences réalisées sur des IN tronquées d'un ou des deux domaines N-ter et C-ter de manière à contrôler que le DIBA-1 n'agit que sur le domaine N-ter montrent, de manière surprenante, que toutes les constructions sont sensibles au DIBA-1. Ceci implique donc que le DIBA-1 possède un deuxième site d'interaction et qu'il est situé au niveau du CC. Trois cystéines sont potentiellement candidates comme interagant du DIBA-1 : les Cys 56, 65 et 130. Nous avons rapidement écarté la C130 car d'une part, la mutation de ce résidu modifie la stabilité de l'IN entière (Bischerour *et al.*, 2003) et d'autre part des données préliminaires de spectrométrie de masse montrent que la cystéine 130 n'interagit pas avec le DIBA de manière stable. Ces résultats préliminaires indiquent également que l'une des deux autres Cys (C55, C56) est touchée. Cependant, avec le protocole utilisé, il est impossible de les différencier, nous avons donc réalisé une mutagenèse dirigée contre ces deux cystéines (C56 et C65). Les tests d'activité en présence de DIBA-1 montrent que les mutants doubles et triples contenant la mutation C56S sont significativement moins sensibles au DIBA-1 dans le contexte  $Mn^{2+}$  alors qu'ils présentent une sensibilité équivalente au simple mutant C280S utilisé comme contrôle en présence de  $Mg^{2+}$ . La mutation C65S montre des sensibilités au DIBA-1 équivalente au contrôle quelque soit le contexte cationique. Ces résultats suggèrent que le DIBA-1 interagit avec la C56 et que son action n'est visible que dans un contexte  $Mn^{2+}$ . Ces observations impliquent donc que dans un contexte  $Mg^{2+}$  l'action inhibitrice du

DIBA-1 se fait principalement au niveau de la C40 du domaine N-ter avec une perte de la coopérativité alors qu'en présence de  $Mn^{2+}$  c'est l'interaction avec la C56 qui est responsable de l'inhibition de l'activité de 3'-processing. Ces résultats confirment notre hypothèse que la structuration correcte du N-ter apparaît comme primordiale pour une multimérisation correcte de l'IN sur l'ADN afin qu'elle catalyse le 3'-processing en présence de  $Mg^{2+}$ . Dans le contexte  $Mn^{2+}$ , les différents résultats nous permettent de penser que la multimérisation coopérative de l'IN n'est pas essentielle pour qu'elle soit catalytiquement active ou qu'un mode d'assemblage différent du multimère actif survient en présence de  $Mn^{2+}$  par rapport au  $Mg^{2+}$  pour lequel la multimérisation suit un processus coopératif médié par le N-ter.

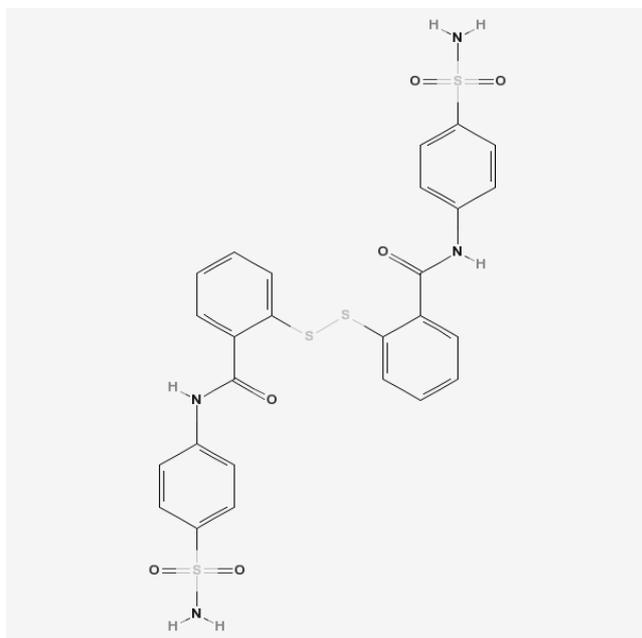
Les études menées sur les protéines tronquées ont également permis de mettre en évidence une affinité différentielle des intégrases tronquées de leur domaine C-ter (IN<sup>55-212</sup> et IN<sup>1-212</sup>) pour l'ADN viral selon le cation divalent présent. Dans un contexte  $Mn^{2+}$ , les intégrases tronquées possèdent toutes des affinités comparables. En présence de  $Mg^{2+}$ , l'absence du domaine C-ter diminue l'affinité pour l'ADN viral des IN ainsi tronquées. Ces données suggèrent l'existence de contacts spécifiques entre l'ADN viral et le domaine C-ter en présence de  $Mg^{2+}$ . Ceci serait symptomatique d'une organisation conformationnelle de l'IN différente selon le cation divalent présent au niveau de son site actif. Cette observation est corroborée d'une part par les profils de spécificité différents vis-à-vis de la séquence de l'ADN en fonction de la nature du cofacteur cationique (Esposito *et al.*, 1998) et d'autre part par des profils de digestion et une reconnaissance par des anticorps différents de l'IN selon la présence ou non de cation divalent au site actif (Asante-Appiah *et al.*, 1998).

Du point de vue pharmacologique, le mode d'inhibition non compétitif de l'ADN du DIBA-1 montre qu'une molécule agissant sur les états multimériques de l'IN peut présenter des  $IC_{50}$  de l'ordre de la dizaine de nanomolaires. Cependant, le DIBA-1 ne présente, en elle-même, qu'un intérêt pharmacologique limité de part sa faible spécificité d'action puisqu'elle agit également sur un récepteur aux oestrogènes humain (Wang *et al.*, 2004), et sur la NCp7 du VIH-1 ((Turpin *et al.*, 1996); (Huang *et al.*, 1998); (Berthoux *et al.*, 1999)). Ceci étant dit, ce mode d'inhibition de type allostérique présente un intérêt majeur dans le cadre d'une volonté de diversifier les cibles et les modes d'actions des antirétroviraux de manière à éviter les résistances croisées entre composés actifs.



### III.E. Figures et Tableaux

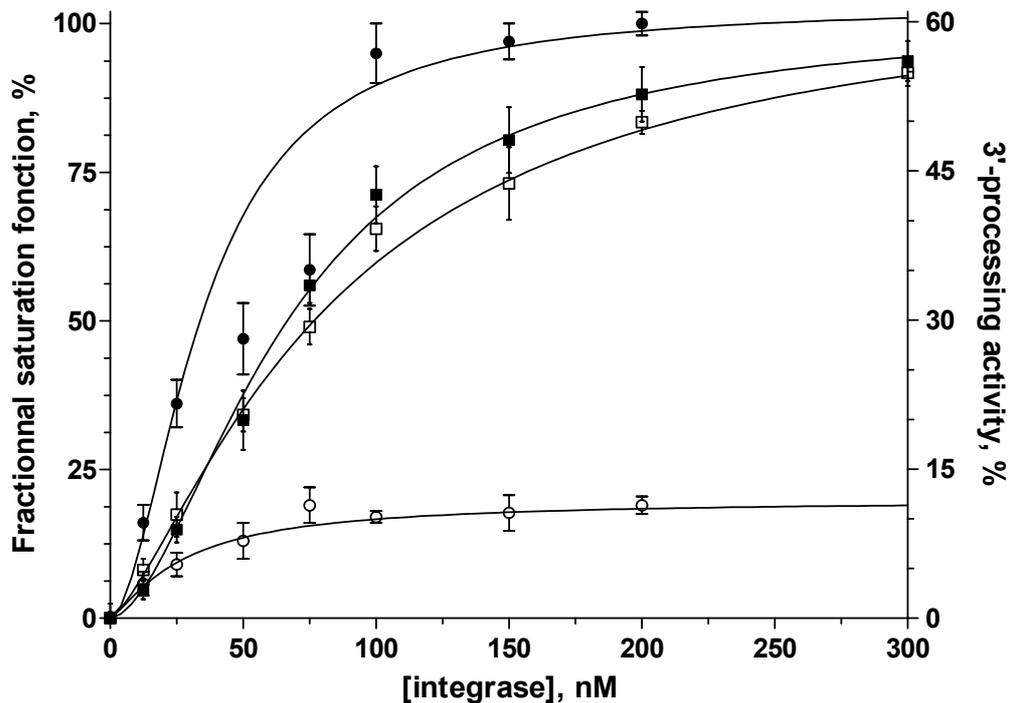
**Figure 1 : Structure du DIBA-1**



**Figure 2: Inhibition des activités de 3'-processing et de désintégration par le DIBA-1. (Cf. ci-contre)**

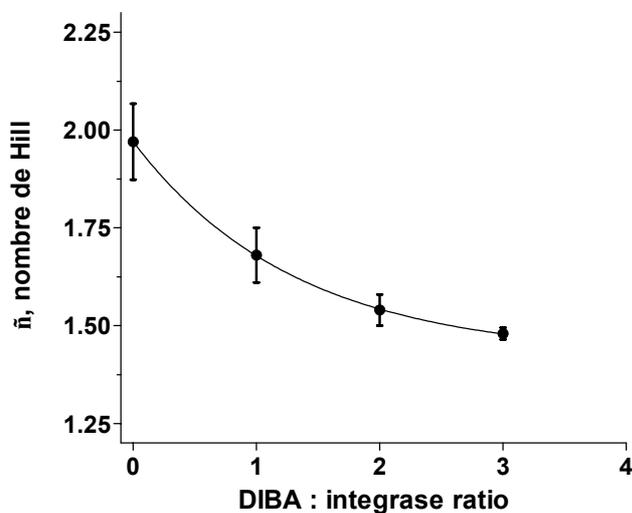
Les réactions de 3'-processing (A) ou désintégration (B) sont réalisées en présence de concentrations croissantes de DIBA comme cela est décrit dans le Matériels et Méthodes, en présence de 10 mM  $Mg^{2+}$  (gauche) ou 10 mM  $Mn^{2+}$  (droite). Les concentrations de l'IN entière et de l'ADN substrat sont respectivement de 50 nM et de 1 nM (sauf pour le puit 1: sans IN). Les concentrations de DIBA-1 sont de 0 (puit 12), 1.9 nM (puit 11), 3.8 nM (puit 10), 7.7 nM (puit 9), 15.4 nM (puit 8), 30.8 nM (puit 7), 61.5 nM (puit 6), 123 nM (puit 5), 370 nM (puit 4), 1111 nM (puit 3) or 3333 nM (puit 2). Les activités relatives de 3'-processing (C) ou de désintégration (D) (avec du  $Mg^{2+}$  - carrés noirs – ou du  $Mn^{2+}$  - cercles blancs – comme cofacteur cationique) sont tracées en fonction de la concentration en DIBA-1. En présence de  $Mg^{2+}$ , les valeurs calculées d'IC<sub>50</sub> sont respectivement de 13 et 11 nM pour les réactions de 3'-processing et de désintégration. Les valeurs correspondantes en présence de  $Mn^{2+}$  sont respectivement de 118 et 123 nM. (E) Effets d'un agent réducteur sur l'activité inhibitrice du DIBA-1. La réaction de 3'-processing est réalisée en utilisant l'IN entière (100 nM), l'ADN substrat (1nM) et du  $Mg^{2+}$  (10 mM) comme cofacteur (puits 2-9). Puit 3, + DIBA (200 nM); puit 4, + DTT; puit 5, +  $\beta$ -mercaptoéthanol; puits 6 et 8, l'IN fut préincubée avec le DIBA pendant 10 min à température ambiante avant l'ajout de l'agent réducteur et de l'ADN (puit 6, DTT; puit 8,  $\beta$ -mercaptoéthanol); puits 7 et 9, l'IN est préincubée en présence de l'agent réducteur (puit 7, DTT; puit 9,  $\beta$ -mercaptoéthanol) pendant 10 min à température ambiante avant l'ajout du DIBA (200 nM) et de l'ADN. Les concentrations de DTT et de  $\beta$ -mercaptoéthanol sont de 4 mM. Puit 1, ADN seul.

**Figure 3: le DIBA ne prévient pas la formation des complexes IN-ADN.**

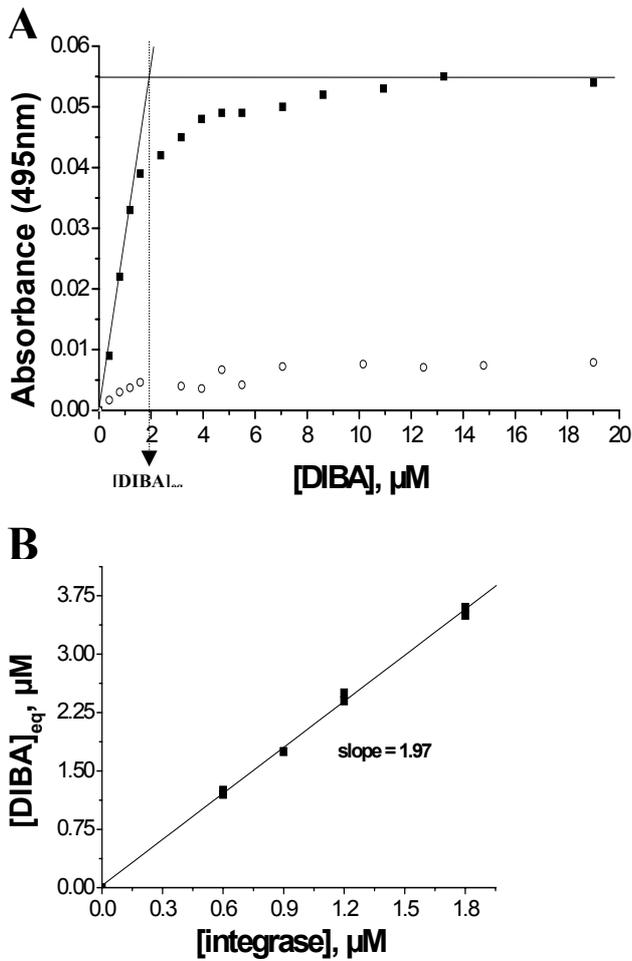


Des concentrations croissantes d'IN furent incubées avec l'ADN substrat (1 nM) dans un tampon Tris (20 mM, pH 7.2) contenant 15% DMSO (v/v), 50 mM de NaCl et 10 mM de  $Mg^{2+}$ . Sur l'axe à gauche, l'étape de fixation de l'IN sur l'ADN fut mesurée par anisotropie de fluorescence statique comme cela est décrit dans le Matériels et Méthodes (carrés noirs, pas de DIBA ; carrés blancs, ratio [DIBA]:[IN] = 2:1). Sur l'axe de droite, l'activité de 3'-processing est mesurée pour chacun des mélanges IN / ADN (cercles noirs, pas de DIBA ; cercles blancs, ratio [DIBA]:[IN] = 2:1).

**Figure 4: nombre Hill en fonction du ratio DIBA:IN.**



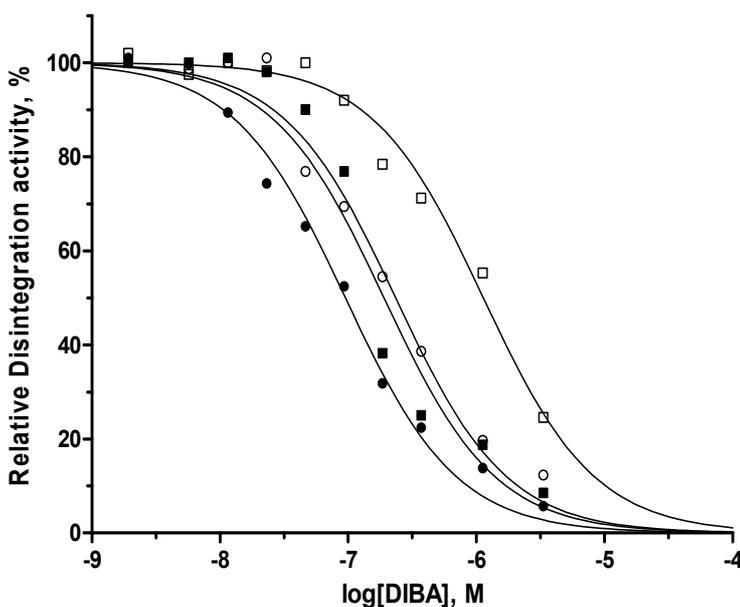
Les isothermes de fixation à l'ADN sont enregistrées comme cela est décrit dans la légende de la figure 2.



**Figure 5: le DIBA induit l'éjection du zinc de l'IN.**

(A) L'éjection du zinc est mesurée par absorbance optique à 495 nm sur des échantillons contenant de l'IN entière (1 μM) et du PAR (10<sup>-4</sup> M) dans un tampon Tris (20 mM, pH 7.0) contenant 20 % DMSO (v/v) en l'absence d'agent réducteur (carrés noirs) ou en présence de 4 mM de β-mercaptoéthanol (cercles blancs). La concentration de DIBA pour éjecter la totalité du zinc – [DIBA]<sub>eq</sub> – est estimé sur le graphe. Cette valeur est tracée en fonction de la concentration initiale d'IN (B). La pente de la régression linéaire indique que l'éjection du zinc coïncide avec la fixation de deux molécules de DIBA par protomère d'IN.

**Figure 6: Effets du DIBA sur l'IN entière et sur les mutants tronqués.**



Les réactions de désintégration sont réalisées en présence de 10 mM de Mn<sup>2+</sup> comme cela est décrit dans le Matériels et Méthodes. Les activités relatives de désintégration sont tracées en fonction de la concentration en DIBA (cercles noirs, IN entière; cercles blancs, ΔN; carrés blancs, ΔC; carrés noirs, CC). Les valeurs d'IC<sub>50</sub> calculées sont respectivement de 95, 234, 1135 et 191 nM pour l'IN entière, la ΔN, la ΔC et le CC. Les concentrations en protéines et en substrat sont respectivement de 50 nM et de 1 nM.

**Tableau 1 : Nombres de Hill pour la fixation de l'IN entière sur l'ADN**

Mg <sup>2+</sup>		Mn <sup>2+</sup>	
DIBA -	DIBA + <sup>a</sup>	DIBA -	DIBA + <sup>a</sup>
1.97 ± 0.10	1.47 ± 0.02	1.29 ± 0.09	1.20 ± 0.10

<sup>a</sup> correspond à un ratio DIBA:IN de 3:1

**Table 2 : Valeurs de K<sub>d,app</sub> (nM)**

IN	Mg <sup>2+</sup>		Mn <sup>2+</sup>	
	DIBA -	DIBA + <sup>a</sup>	DIBA -	DIBA + <sup>a</sup>
Entière	65 ± 7	65 ± 9	51 ± 7	53 ± 5
ΔN	45 ± 7	45 ± 6	28 ± 3	29 ± 6
ΔC	531 ± 188	652 ± 214	60 ± 4	64 ± 8
CC	233 ± 98	279 ± 34	53 ± 13	61 ± 10

<sup>a</sup> correspond à un ratio DIBA:IN de 3:1

**Table 3 : les valeurs d'IC<sub>50</sub> caractérisant l'inhibition de l'IN par le DIBA-1 obtenue avec les différents mutants des cystéines**

Cofacteur	IC <sub>50</sub> , nM			
	IN <sub>C280S</sub>	IN <sub>C65S,C280S</sub>	IN <sub>C56S,C280S</sub>	IN <sub>C56S,C65S,C280S</sub>
Mg <sup>2+</sup>	17 ± 3	12 ± 2	19 ± 4	24 ± 5
Mn <sup>2+</sup>	76 ± 16	63 ± 13	1155 ± 547	566 ± 186

**CONCLUSION ET**  
**DISCUSSION GENERALE**

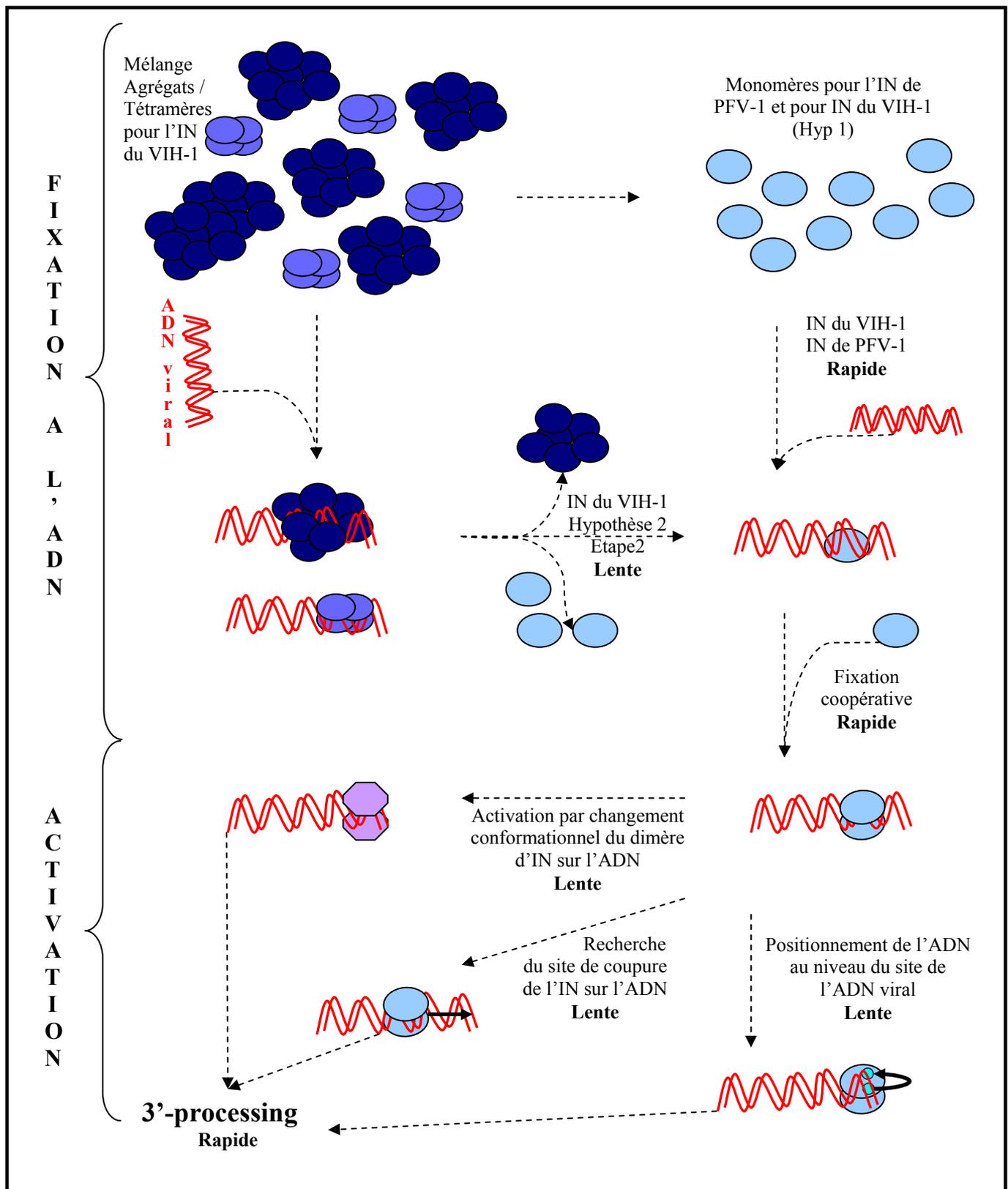
Toutes les études menées au cours de cette thèse ont mis en évidence certaines propriétés biochimiques et catalytiques caractérisant l'IN du VIH-1 et plus généralement les intégrases rétrovirales. Ces études se révèlent primordiales pour les études à venir, qu'elles soient structurales ou pharmacologiques. Nous avons étudié deux IN rétrovirales, celle d'un lentivirus, le VIH-1, et celle d'un spumavirus, le PFV-1. Ces deux IN exercent le même type de fonction *in vivo* et leurs activités *in vitro* sont comparables même si leur profil de solubilité est différent. Au cours de ce travail, nous nous sommes attachés à étudier d'une part les interactions IN / ADN viral et d'une part les paramètres structuraux et cinétiques qui caractérisent la première étape du processus d'intégration : le 3'-processing. Cette étape consiste en une coupure endonucléolytique d'un dinucléotide –GT, pour l'IN du VIH-1, et -AT pour celle de PFV-1, qui suit la séquence consensus CA conservée chez tous les rétrovirus.

Ces deux enzymes ont une fonction et un mécanisme catalytique qui semblent être identiques. Malgré une organisation en trois domaines avec une triade catalytique identique positionnée au niveau du CC, ces deux enzymes possèdent une homologie de séquence de seulement 15%. De ce fait, ces deux IN présentent un certain nombre de différences non négligeables, comme en particulier leur comportement lorsqu'elles sont seules en solution. Au cours de ce travail, nous avons montré que l'IN de PFV-1 présente une transition monomère / dimère avec une demi-transition pour des valeurs de l'ordre de 20 à 30  $\mu\text{M}$  (détermination par TFA et FCS). Par opposition, l'IN du VIH-1 est sous forme agrégé dès que sa concentration s'élève au dessus de 200 nM, au dessous de cette valeur seuil, elle est majoritairement sous forme tétramérique (Deprez *et al.*, 2000). La présence de détergent permet d'améliorer sa solubilité jusqu'à des concentrations micromolaires. Toutefois, dans ces conditions, l'IN du VIH-1 est agrégée au delà de 2  $\mu\text{M}$ . Parallèlement à notre étude de l'IN de PFV-1 par FCS, nous avons mené les mêmes expériences avec l'IN du VIH-1. Mais pour cette dernière, les fluctuations de l'intensité de fluorescence révélaient des pics très intenses de fluorescence, témoins de la présence d'agrégats pour des concentrations de l'ordre du nanomolaire. Pour l'IN de PFV-1, ces pics d'intensité de fluorescence étaient absents ou très peu nombreux jusqu'aux concentrations testées de 200 / 250  $\mu\text{M}$ . L'IN de PFV-1 présente donc un profil de solubilité beaucoup plus clément que celui de l'IN du VIH-1. Les déterminants structuraux d'une telle différence de solubilité reste peu compris. En effet, les profils d'hydrophobicité de ces deux protéines ne présente aucune différence significative. Des constructions de chimères entre ces deux IN montrent que le domaine N-ter pourrait être responsable de la faible

solubilité de l'IN du VIH-1 (Lee *et al.*, 2005a). Cependant, les différentes mutations ponctuelles introduites dans la séquence de l'IN du VIH-1 pour augmenter sa solubilité comme les mutations F185K (Jenkins *et al.*, 1996), W131D, F139D et C280S ((Goldgur *et al.*, 1998); (Chen *et al.*, 2000a)) ne sont pas positionnées dans ce domaine.

Lorsqu'elles sont en présence de leur substrat respectif, les deux IN présentent des profils catalytiques qualitativement comparables. Même si leurs affinités pour leurs ADN viraux respectifs présentent des différences significatives, un  $K_{d,app}$  de 40 nM pour l'IN du VIH-1 et de 130 nM pour celle de PFV-1, nous avons pu montrer par TFA que les deux IN catalysent le 3'-processing sous une forme multimérique identique : un dimère. Cette catalyse est caractérisée par une cinétique de réaction très lente quelque soit l'IN considérée. D'ailleurs, les profils de solubilité des deux IN en présence d'ADN se rapprochent alors qu'ils sont très différents en solution. En effet, les deux IN agrègent sur l'ADN lorsque le ratio IN : ADN est plus élevé que le ratio correspondant à la formation du dimère d'IN. Nous pensons que la construction de ce dimère compétent pour le 3'-processing suit un schéma identique pour les deux IN (cf. figure 53). Comme le montre les expériences de cross-link résolu en temps (Faure *et al.*, 2005) et la coopérativité de fixation à l'ADN d'ordre 2 que nous avons trouvé pour les deux INs en présence de  $Mg^{2+}$ , la fixation des deux monomères se fait de manière séquentielle et coopérative. Il y a dans un premier temps formation d'un complexe Monomère d'IN / ADN, puis ce complexe recrute de manière facilitée un deuxième monomère de façon à former un dimère d'IN sur l'ADN substrat.

Il faut noter que l'ADN viral favorise les contacts IN / IN et induit une modification de l'état multimérique de l'IN. L'IN de PFV-1 est trouvée majoritairement sous forme dimérique sur son ADN substrat pour des concentrations de l'ordre de 400 nM pour lesquelles elle est monomérique lorsqu'elle est seule en solution. L'IN du VIH-1, trouvée majoritairement sous forme tétramérique en solution pour des concentrations inférieures à 200 nM (Deprez *et al.*, 2000), se dissocie en présence d'ADN viral pour reformer un équilibre monomère / dimère fixé sur l'ADN (Deprez *et al.*, 2001). La différence de solubilité entre les deux IN implique des cinétiques de fixation sur l'ADN différentes. L'IN de PFV-1 peut former rapidement des complexes monomères et dimères d'IN / ADN par simple recrutement des monomères déjà présents majoritairement dans le mélange réactionnel (Delelis *et al.*, 2008). Contrairement à elle, l'IN du VIH-1 est présente en solution majoritairement sous forme de tétramères ou d'agrégats. De ce fait, la fixation directe d'un monomère d'IN sur l'ADN, à la manière de l'IN de PFV-1, ne se réalise pas aussi rapidement (Delelis *et al.*, 2008). En effet, la fixation de

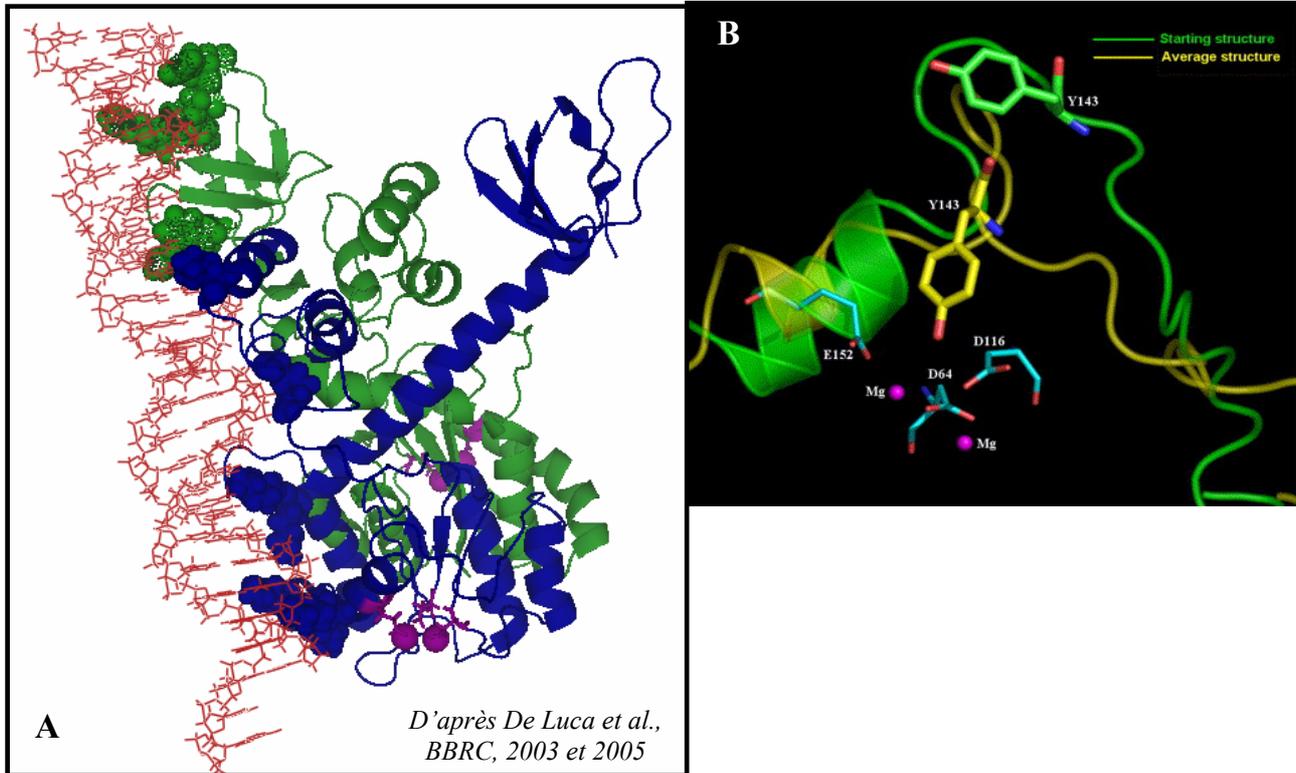


*Figure 53 : Modèle de la formation du complexe dimère d'IN / ADN compétent pour le 3'-processing*

l'IN du VIH-1 sur l'ADN suit une cinétique lente de  $k_{on} \approx 0.23 \text{ min}^{-1}$ . Deux hypothèses permettent d'expliquer cette cinétique lente. (i) Les agrégats et les tétramères d'IN sont en équilibre avec des formes monomériques minoritaires, qui, en se fixant sur l'ADN, vont permettre un déplacement de l'équilibre vers une dissociation lente des formes de hauts poids moléculaires en monomères. Ces monomères pourront alors interagir avec l'ADN. L'autre hypothèse serait que (ii) les agrégats ou les tétramères peuvent interagir directement avec l'ADN. L'interaction induit ensuite, par des changements conformationnels, une dissociation de la forme de haut poids moléculaire laissant alors un monomère sur l'ADN (Deprez *et al.*, 2001) (cf. figure 53).

Nous avons montré qu'en présence de  $Mg^{2+}$ , la fixation du deuxième monomère d'IN sur l'ADN se fait de manière coopérative ( $\tilde{n} = 2$ ) alors qu'en présence de  $Mn^{2+}$ , ce nombre de Hill n'est plus que de 1,1. Des études comparatives d'effet de drogues selon le cofacteur cationique ((Grobler *et al.*, 2002); (Zhang *et al.*, 2003); (Marchand *et al.*, 2003)) mettent en avant une structure du site actif probablement plus flexible en présence de  $Mn^{2+}$  qu'en présence de  $Mg^{2+}$ . Ceci peut s'expliquer par une sphère de coordination du  $Mn^{2+}$  plus flexible que celle du  $Mg^{2+}$  (Charles W. Bock *et al.*, 1999). Différentes études montrent également que la présence d'un cofacteur cationique au site actif induit des changements conformationnels de l'IN dits à longue distance, *i.e.* qui ne sont pas situés dans l'environnement proche du site actif ((Asante-Appiah and Skalka, 1997); (Asante-Appiah *et al.*, 1998)). Ces différents éléments permettent de penser que le cofacteur cationique présent ( $Mg^{2+}$  ou  $Mn^{2+}$ ) peut, par des modifications d'ordre conformationnelles, influencer sur le processus de multimérisation de l'IN sur l'ADN.

Les expériences que nous avons pu mener ont permis d'apporter un certain nombre de réponses concernant les domaines impliqués dans ces deux processus de multimérisation de l'IN. Nous avons montré que le domaine N-ter, impliqué dans le processus de multimérisation de l'IN du VIH-1 lorsqu'elle est seule en solution (Lee *et al.*, 1997), joue également un rôle clé dans la multimérisation coopérative  $Mg^{2+}$ -dépendante. En effet, la déstructuration de ce domaine par éjection de son atome de Zinc induit une perte de la multimérisation coopérative. Les mêmes expériences menées en présence de  $Mn^{2+}$  montrent d'une part que la fixation de l'IN à l'ADN est peu ou pas coopérative et d'autre part qu'elle n'est pas sensible à la structuration du domaine N-terminal. Il n'est d'ailleurs pas exclu par nos expériences que l'IN



**Figure 54 : Modèle du dimère d'IN en présence d'ADN viral traité**

**A** : Modèle d'interaction dimère d'IN (monomère A : vert, B : bleu) / ADN (rouge). En boule, les résidus interagissant avec l'ADN (C-ter monomère A : R231, S230, R263, K264, W243, K244, N-ter et CC monomère B : R20, K14, K188, K186, K160, K159, K156). En magenta, la triade catalytique et les deux  $Mg^{2+}$  de chaque site actif. *D'après (De Luca et al., 2003)*

**B** : Dynamique moléculaire sur ce modèle au niveau de la boucle flexible et du positionnement de la Tyr 143. Mouvement de la boucle (de vert à jaune) observé uniquement que pour le site catalytique du monomère A qui réalise la catalyse. Au niveau du site actif : positionnement de deux cations divalent  $Mg^{2+}$  en accord avec la structure d'ASV. *D'après (De Luca et al., 2005)*

puisse catalyser le 3'-processing  $Mn^{2+}$ -dépendant sous forme monomérique. Ceci pourrait corroborer le fait que la déstructuration du domaine N-terminal influe sur l'activité de 3'-processing en présence de  $Mg^{2+}$ , mais pas en présence de  $Mn^{2+}$  ((Zheng *et al.*, 1996); (Lee *et al.*, 1997)). L'ensemble des données structurales et nos expériences montrent que le CC et le domaine N-ter jouent un rôle primordial pour l'assemblage du dimère actif dans un contexte  $Mg^{2+}$ , reconnu comme étant le seul contexte propice à l'activité spécifique.

Si l'on observe le positionnement du domaine N-ter de l'IN du VIH-1, que se soit au niveau de la structure du double domaine N-Ter / CC (Wang *et al.*, 2001) (cf. figure 37), au niveau de la structure hypothétique du dimère d'IN du VIH-1 (Wang *et al.*, 2001) (cf. figure 39), ou au niveau du modèle structural du dimère d'IN en présence d'ADN viral traité (De Luca *et al.*, 2003) (cf. figure 54 A), ce domaine semble occuper une place importante (i) au sein de l'interface de dimérisation de l'IN. En effet, ce domaine forme d'une part des interactions N-ter / N-ter, et d'autre part, chacun d'eux interagit également avec les CC (ii) dans les interactions IN / ADN, puisque plusieurs résidus, par des études de cross-link et de docking, forment avec le squelette sucre-phosphate de l'ADN des interactions ioniques (K14, R20) mais également un certain nombre d'interactions hydrophobes avec des intercalations partielles possibles des chaînes latérales entre les bases nucléotidiques (F1, L2, W19, A21, M22, A23, F26 et P30) (De Luca *et al.*, 2003), (Zhao *et al.*, 2008). Tout changement conformationnel de ce domaine, que se soit par l'éjection du Zinc, ou par un changement induit à longue distance par le cofacteur cationique ne modifiera pas l'affinité apparente de l'IN pour son ADN substrat puisque le CC et le C-ter interagissent aussi avec l'ADN viral. Par contre, il va modifier les surfaces d'interaction IN / ADN et IN / IN. Ces changements peuvent expliquer ces deux processus différents de dimérisation de l'IN en présence d'ADN.

L'ADN, en favorisant les interactions IN / IN, va également favoriser la formation de complexes IN / ADN de haut poids moléculaire. En effet, lorsque le ratio IN : ADN est augmenté, les temps de corrélations passent rapidement de 30 / 35 ns (temps de corrélation du dimère d'IN du VIH-1 sur l'ADN) pour un ratio de 1 : 40 à plus de 90 ns à partir d'un ratio de 1 : 100 qui correspond à la présence d'une majorité d'agrégats d'IN sur l'ADN. La résolution des temps de corrélation, auxquels nous avons accès, ne nous permet pas de dire si avant d'atteindre l'état d'agrégat sur l'ADN, l'état multimérique de l'IN passe par un tétramère compétent pour le 3'-processing. Il existe actuellement deux études contradictoires concernant l'activité de 3'-processing du tétramère : des tests d'activité en présence de tétramères d'IN

stabilisés par cross-link montrent que cette forme n'est pas compétente pour le 3'-processing (Faure *et al.*, 2005). Une deuxième étude montre que l'IN du VIH-1 peut également catalyser la réaction de 3'-processing sous un état multimérique d'ordre supérieur à deux, et que cet état est probablement un tétramère (Delelis *et al.*, 2007). L'IN d'ASV peut également catalyser le 3'-processing sous forme dimérique et tétramérique avec pour chacune des deux formes une cinétique différente : les auteurs trouvent pour le dimère d'IN d'ASV une cinétique du même ordre que celle que nous avons trouvée pour les IN du VIH-1 et de PFV-1 ((Guiot *et al.*, 2006); (Smolov *et al.*, 2006); (Delelis *et al.*, 2008)) et pour son tétramère une cinétique 100 fois plus rapide (Bosserman *et al.*, 2007). Il est à noter que nos cinétiques de 3'-processing n'ont jamais permis de mettre en évidence un comportement biphasique. A ce jour, la recherche sur les relations entre les états multimériques de l'IN du VIH-1 et les activités catalytiques résultantes privilégie, malgré ces deux études, l'hypothèse d'un 3'-processing catalysé par un dimère (Guiot *et al.*, 2006) principalement stabilisé au niveau de l'interface CC / CC, mais dont l'assemblage coopératif est certainement médié par le domaine N-terminal.

Il semble établi que la deuxième étape du processus d'intégration (l'intégration concertée) est catalysée par un tétramère d'IN ((Faure *et al.*, 2005); (Li *et al.*, 2005); (Li *et al.*, 2006)). Ce dernier pourrait être formé par la dimérisation de deux dimères d'IN. Des études montrent que cette multimérisation d'un ordre supérieur serait plutôt médiée par le CC et le domaine C-ter. En effet, plusieurs résidus du CC semblent impliqués dans ce processus comme les résidus F185 ((Podtelezhnikov *et al.*, 2003) ou la boucle 188-194 ((Wang *et al.*, 2001); (Bosserman *et al.*, 2007)) dont les mutations induisent une diminution des activités attribuées aux tétramères d'IN. Certains résidus du CC d'autres IN rétrovirales comme l'IN d'ALSV furent également montrés comme primordiaux pour l'activité tétramère dépendante comme I88 et L135, qui correspondent à I84 et C130 pour l'IN du VIH-1, alors que leurs activités dimère dépendante n'est pas touchée (Moreau *et al.*, 2004). L'étude de mutants de l'IN du VIH-1 comme l'IN W235A et W235F (Li *et al.*, 2005), ainsi que les données structurales du double domaine IN<sup>55-288</sup>, montrent que les deux domaine C-ter du dimère de la maille cristallographique, de part leur écart de 55 Å, n'interagissent pas entre eux, mais ils forment des interactions C-ter / C-ter avec les dimères des autres mailles au niveau des interfaces C et D. Il faut d'ailleurs noter que le seul contact non hydrophobe C-ter / C-ter de l'interface C est une liaison hydrogène entre R224N $\eta$ 1 d'un C-ter avec l'oxygène du carbonyle de la chaîne principale du Trp 235 de l'autre. Le Trp 235 est également impliqué dans les

interactions hydrophobes C-ter / C-ter de l'interface D qui implique la boucle 242-246 d'un dimère et le brin  $\beta 5$  265-269 d'un autre (Chen *et al.*, 2000a). Ces deux types d'interaction expliquent pourquoi le maintien du cycle benzénique de ce résidu est important pour l'intégration concertée (qui nécessite un tétramère d'IN) et non pour l'activité de 3'-processing (Li *et al.*, 2005). Le domaine C-ter jouerait donc un rôle dans la tétramérisation de l'IN alors qu'il n'aurait qu'un impact très limité dans sa dimérisation. Ces observations ont également été faites pour d'autres IN rétrovirales comme l'IN d'ALSV pour laquelle différents mutants du domaine C-ter (K266A, V239A et V249A) présentent une diminution de l'intégration concertée, reliée à la tétramérisation de l'IN, sans modification de sa capacité à dimériser (Moreau *et al.*, 2003). Cette dimérisation de dimères permet notamment de rapprocher dans l'espace les deux extrémités de l'ADN viral de manière à les insérer à cinq paires de bases d'écart dans l'ADN de la cellule hôte.

La plupart des modèles concernant des complexes IN / ADN sont réalisés sur ce complexe tétramère d'IN du VIH-1 / ADNs viraux / ADN cible responsable de l'intégration concertée. Un certain nombre d'entre eux est calqué sur la structure résolue du complexe transposase Tn5 / ADN, qui est fonctionnellement très proche ((Heuer and Brown, 1998); (Gao *et al.*, 2001); (Podtelezchnikov *et al.*, 2003); (Karki *et al.*, 2004); (Wielens *et al.*, 2005); (Chen *et al.*, 2006), pour les tétramères en présence des deux ADN viraux). Ils possèdent tous une organisation symétrique de manière à réaliser les deux transferts de brin à cinq paires de bases d'écart. Chacun des modèles considèrent également que l'intégration concertée est catalysée par deux sites actifs différents. Par contre, ces modèles diffèrent dans l'organisation géométrique et dans le positionnement des ADN viraux et cibles. Le modèle de Heuer fut le premier publié. Le complexe modélisé est le seul qui comporte quatre dimères d'IN. Les autres sont constitués de deux dimères. Ce premier modèle octamérique donne un positionnement des ADN viraux qui se rapproche de celui proposé par le modèle de Wielens et al : les ADN viraux joue le rôle de pont entre les différents monomères d'IN. En effet, dans ces deux modèles, les ADN viraux interagissent avec le monomère qui a déjà catalysé le 3'-processing par les quelques bases terminales. Les bases plus éloignées vont interagir avec le monomère d'un dimère adjacent. Dans les autres modèles, l'ADN viral interagit avec le monomère catalytiquement actif par le CC et le N-ter. Les bases plus éloignées interagissent alors avec le C-ter du domaine situé en trans- dans le dimère responsable de l'activité de 3'-processing, comme c'est le cas pour le dimère d'IN modélisé par De Luca (De Luca *et al.*, 2003). Le modèle de Gao et al, comme celui de Heuer, ne positionne pas les domaines N-ter.

Son organisation structurale est composée d'un dimère de dimère qui repose sur la structure du double domaine CC - C-ter et sur les données de cross-link entre l'IN et l'ADN viral. Il propose que les dimères interagissent par les domaines N-ter. Le modèle de Podtelezchnikov et al. calque la structure du dimère de dimère sur la structure du complexe dimère de Tn5 transposase / ADN résolue par Davies et al. (Davies *et al.*, 2000). Dans ce modèle, les auteurs proposent que les deux dimères forment une large surface d'interaction et sont positionnés sur la même face de l'ADN cellulaire. Le modèle de Wielens est également basé sur la structure du complexe Tn5 transposase / ADN. Par contre, les deux dimères interagissent uniquement par les domaines N-ter et les deux CC catalytiquement actifs sont situés sur les deux faces opposées de l'ADN cible. Les C-ter des deux monomères actifs et les deux monomères non actifs sont placés de manière à stabiliser l'ensemble. Ces modèles, même s'ils s'appuient sur les mêmes données structurales et biochimiques, montrent tous une organisation symétrique avec une variabilité dans le positionnement des ADN viraux et des domaines C-ter et N-ter. La reconstruction à partir d'images, obtenues au microscope électronique, du complexe tétramérique d'IN en présence d'ADN mimant le produit de l'intégration concertée, jette le trouble quant à la véracité de ces modèles (Ren *et al.*, 2007). En effet, la reconstruction montre une organisation asymétrique du complexe. Elle pourrait être due soit à une réelle asymétrie dans l'organisation tridimensionnelle du complexe IN / ADN viral / ADN cible, soit être la représentation d'une moyenne qui pourrait être relevante d'un large changement d'organisation du complexe. Un tel changement d'organisation tridimensionnelle, qui implique donc une certaine hétérogénéité de structure entre les différents complexes présents en solution, pourrait également expliquer pourquoi les cristaux d'IN du VIH-1 obtenus par Chen et al. présentent une diffraction permettant une résolution de seulement 8 Å (Chen *et al.*, 2000a).

Notre étude des cinétiques de 3'-processing par anisotropie de fluorescence a révélé que ces deux IN rétroviraux présentent des cinétiques de catalyse lentes du même ordre de grandeur :  $0.004 \text{ min}^{-1}$  pour l'IN du VIH-1 (Guiot *et al.*, 2006) et  $0.0028 \text{ min}^{-1}$  pour l'IN de PFV-1. Plusieurs modèles peuvent être envisagés pour expliquer ce phénomène lent (cf figure 53):

- A ce jour, aucune spécificité de reconnaissance de l'ADN n'a été mise en évidence pour l'IN du VIH-1. Nous avons pu montrer que l'IN de PFV-1, même si elle présente une certaine tendance à interagir de manière plus spécifique avec l'extrémité processée de l'ADN viral, interagit également de manière non spécifique avec lui. L'IN pourrait

donc se fixer sur l'ADN viral de manière non spécifique et se déplacer le long de l'ADN jusqu'à son site de coupure spécifique, à la manière des enzymes de restriction EcoRI et EcoRV ((Surby and Reich, 1996); (Jeltsch *et al.*, 1996)). Cependant, ce modèle à lui seul ne permet pas d'expliquer complètement la lenteur des cinétiques observées car les déplacements en une dimension sur l'ADN observés pour différentes enzymes ne sont généralement pas de l'ordre de l'heure, mais plutôt de la sec ou de la msec (Jack *et al.*, 1982). De plus, lors de nos expériences de FRET entre l'IN de PFV-1 et l'extrémité de l'ADN viral processée, nous n'avons pas pu observer de variation du FRET au cours du temps qui aurait pu être symptomatique d'un mouvement directionnel de l'IN vers l'extrémité processable de l'ADN.

- La deuxième hypothèse fait intervenir l'existence de deux sites de fixation de l'ADN au niveau du CC : un pour l'ADN viral et un pour l'ADN cible (Harper *et al.*, 2001). Ceci est corroboré par les modes d'inhibition des InBI et des InSTI. En effet, les InBI sont des compétiteurs de l'ADN viral (Deprez *et al.*, 2004) et les InSTI, inhibiteurs spécifiques du transfert de brin, sont principalement des compétiteurs de l'ADN cible et ne modifient donc pas l'affinité de l'IN pour l'ADN viral (Espeseth *et al.*, 2000). Il faut ajouter à cela que ces deux familles d'inhibiteurs possèdent des spectres de mutations de résistance non recouvrant, ce qui indique que ces drogues agissent au niveau de sites différents. Il faut également noter d'une part qu'un seul site actif est responsable des deux réactions successives (Engelman *et al.*, 1992) et que d'autre part les données expérimentales ou de docking moléculaire indiquent que ces deux sites d'interaction avec l'ADN sont proches ((Marchand *et al.*, 2002); (Perryman and McCammon, 2002)). L'ADN substrat pourrait donc se fixer indifféremment au niveau de chacun de ces deux sites. L'interaction du substrat au niveau du site de l'ADN cible ne permet pas la catalyse du 3'-processing. Par contre, ces complexes peuvent, si le substrat change de site par un mécanisme lent, devenir catalytiquement actif.
- La troisième hypothèse fait intervenir un changement conformationnel lent de l'IN qui passe d'un état inactif à un état catalytiquement actif rendant ainsi possible le 3'-processing. Cette acquisition lente de la compétence peut être due, par exemple, à un changement conformationnel de la boucle flexible comprise entre les résidus G140 et G149 du CC responsable de la coupure (modélisé par De Luca *et al.* (cf. figure 54 B) (De Luca *et al.*, 2005)). Des études de mutagenèses dirigées sur la boucle pourraient confirmer ce modèle. L'activation du dimère pourrait également être due à un changement de position d'un domaine : différentes expériences de coss-link entre l'IN

et l'ADN viral ou l'ADN viral traité montrent par exemple que le résidu S230 n'interagit pas avec les mêmes bases lorsque l'ADN est non- ou pré-traité. De ce fait, le complexe IN / ADN pourrait adopter différentes conformations au cours de l'activité de 3'-processing (Gao *et al.*, 2001).

A ce jour, sur ces trois modèles, les deux derniers semblent donc les plus plausibles.

Nous avons également montré pour les deux IN (VIH-1 et PFV-1) un mécanisme catalytique de type « single turn-over » lors du 3'-processing. Ce mécanisme peut s'expliquer par le caractère séquentiel du processus global d'intégration. L'IN, en réalisant le 3'-processing de l'ADN viral, produit un ADN traité qui est le substrat de la réaction de transfert de brin. Le 3'-processing étant la première étape, le complexe Enzyme / Substrat doit rester en contact de manière à pouvoir réaliser la suite du processus enzymatique. Ce système permet d'optimiser la catalyse du processus complet. Ce type de mécanisme est également utilisé par les transposases comme la Tn5 qui réalisent quatre réactions séquentielles au niveau du même site actif (répétition d'un clivage endonucléolytique et d'un transfert de brin) (Bhasin *et al.*, 1999). Cette faible efficacité, en apparence, n'a pas d'implication biologiquement parlant puisque ces enzymes n'ont besoin de réaliser qu'un seul acte intégratif complet pour être considérées comme efficaces. Au contraire, un trop grand nombre d'actes intégratifs pourraient avoir des conséquences délétères sur la survie cellulaire et par conséquent sur la multiplication lentivirale.

Ainsi, cette étude montre d'une part l'implication du domaine N-ter de l'IN et du cofacteur cationique dans la formation coopérative d'un complexe dimère d'IN / ADN catalytiquement compétent pour le 3'-processing de l'IN. Ceci étant dit, même si nous avons plusieurs hypothèses, nous n'avons pas réussi à expliquer les raisons de cette cinétique lente.

Nous avons montré qu'un inhibiteur allostérique modifiant cette coopérativité pouvait présenter des  $IC_{50}$  de l'ordre du nanomolaire compatibles avec des essais de développement *ex* puis *in vivo*. Ce type d'inhibiteur est particulièrement intéressant puisqu'ils agissent au niveau de sites différents de ceux ciblés par les InBI et les InSTI.

D'autre part, nous avons montré que L'IN de PFV-1 de part son profil de solubilité et par ses propriétés enzymatiques semblables à celles de l'IN du VIH-1 est un bon candidat pour tenter de résoudre la structure d'une IN rétrovirale entière. Il reste à définir les conditions de cristallisation pour lesquelles l'IN est dans une conformation la plus stable possible, notamment si elle est en présence d'un cofacteur, de son substrat et/ou d'un inhibiteur.

# **ANNEXES**

## ***Annexe 1 : Composition des tampons utilisés lors de la purification de l'IN***

### **1L de Tampon A:**

58,44g de NaCl (1M final)

50mL de TrisHCl pH 8 (50mM final)

380 $\mu$ L de betamercaptoéthanol à 99% (4mM final)

Compléter à 1L et filtrer avec un filtre de 0.40 $\mu$ m de diamètre

### **50mL d'IMI 2M**

10g d'Imidazole hydrochloride

50mL de Tampon A

Vortexer puis filtrer

### **50mL de tampon de Resuspension**

0,5mL d'IMI 2M

25mL de Tampon A

Dissoudre une pastille d'antiprotéase complète **EDTA free** (Roche) puis filtrer

### **20mL de Tampon d'élution**

10mL de la solution d'IMI 2M

10mL de tampon A

1 $\mu$ L de ZnSO<sub>4</sub> 1M (50mM final)

Filtrer

### **2L Tampon de dialyse**

116.8g de NaCl (1M final)

800mL H<sub>2</sub>O

40mL HEPES pH7 (20mM final)

200mL glycerol (10% final)

766 $\mu$ L  $\beta$ -mercaptoéthanol (4mM final)

100 $\mu$ L ZnSO<sub>4</sub> 1M compléter avec H<sub>2</sub>O à 2L

## ***Annexe 2 : Intégrase et intégration, activités biochimiques de l'intégrase du VIH-1***

Cette revue fait le point sur les activités biochimiques de l'IN du VIH-1 mises en évidence *in vivo* et *in vitro*.

### **Résumé**

L'intégration de l'ADN rétroviral est une étape obligatoire dans le cycle de réplication des rétrovirus. L'intégrase, la troisième enzyme codée par le génome rétroviral, catalyse cette réaction. Cette dernière se déroule en deux étapes séquentielles avec une séparation spatiale comme temporaire. Le 3'processing, la première, correspond à la maturation de l'ADN viral nouvellement synthétisé par la transcriptase inverse par coupure endonucléolytique d'un dinucléotide GT au niveau d'une séquence conservée CAGT. L'ADN est alors prêt pour être intégré de manière concertée dans l'ADN cible, la deuxième étape. L'intégration concertée à lieu dans le noyau alors que le 3'processing est réalisé dans le cytoplasme cellulaire. Une nouvelle réaction *in vitro* a été mise en évidence, chez l'IN du VIH-1 et PFV-1, au laboratoire. Elle consiste en une coupure endonucléolytique d'une séquence palindromique mimant la jonction LTR-LTR présente au niveau des cercles 2-LTR. Cette revue met l'accent sur la capacité de l'IN à se lier sur les séquences symétriques, ce qui pourrait donner des perspectives quant au ciblage de séquences spécifiques par l'IN lors de thérapies géniques.

### **Manuscrit**

Cette revue a été soumise dans « Retrovirology ».

REVIEW

*INTEGRASE AND INTEGRATION*

Biochemical activities of HIV-1 Integrase

Olivier Delelis<sup>\*1</sup>, Kevin Carayon<sup>1</sup>, Ali Saïb<sup>2</sup>, Eric Deprez<sup>1</sup> and Jean-François Mouscadet<sup>1</sup>.

<sup>1</sup>LBPA, CNRS UMR8113, Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, 94235 Cachan Cedex, France. <sup>2</sup>CNRS, UMR7151, Hôpital Saint-Louis, 1 Avenue Claude Vellefaux, 75475 Paris Cedex 10, France.

\* Corresponding author

Email: Olivier Delelis – delelis@lbpa.ens-cachan.fr

## ABSTRACT

Integration of retroviral DNA is an obligatory step of retrovirus replication because proviral DNA is the template for productive infection. Integrase, a retroviral enzyme, catalyses integration. The process of integration can be divided into two sequential reactions. The first one, named 3'-processing, corresponds to a specific endonucleolytic reaction which prepares the viral DNA extremities to be competent for the subsequent covalent insertion, named strand transfer, into the host cell genome by a trans-esterification reaction. Recently, a novel specific activity of the full length integrase was reported, *in vitro*, by our group for two retroviral integrases (HIV-1 and PFV-1). This activity of internal cleavage occurs at a specific palindromic sequence mimicking the LTR-LTR junction described into the 2-LTR circles which are peculiar viral DNA forms found during viral infection. Moreover, recent studies demonstrated the existence of a weak palindromic consensus found at the integration sites. Taken together, these data underline the propensity of retroviral integrases for binding symmetrical sequences and give perspectives for targeting specific sequences used for gene therapy.

## INTRODUCTION

The human immunodeficiency virus is the causal agent of AIDS. AIDS morbidity and mortality have led to efforts to identify effective inhibitors of the replication of this virus. Viral replication is driven by a molecular motor consisting of the three viral enzymes: the reverse transcriptase, protease and integrase (IN). The genomic RNA of the virus is used to produce a copy of viral DNA by reverse transcription, and the last of these enzymes, integrase, catalyses the covalent insertion of this DNA into the chromosomes of the infected cells. Once integrated, the provirus persists in the host cell and serves as a template for the transcription of viral genes and replication of the viral genome, leading to the production of new viruses. Integrase possesses two major catalytic activities: an endonucleolytic cleavage at each 3'-OH extremities of the viral genome, named 3'-processing, and a strand transfer reaction leading to the insertion of the processed viral DNA into the target DNA by a trans-esterification mechanism. These catalytic functions of the integrase are essential for the overall integration process and have thus been the object of intensive pharmacological research. Since the end of the 1990s, several inhibitors with genuine antiviral activity have been identified and developed. Two of these compounds — MK-0518 or raltegravir and GS9137 or elvitegravir — have shown great promise and should ensure that integrase inhibitors rapidly become an important class in the arsenal of antiretroviral drugs (ARVs) available [1]. In addition to 3'-processing and strand transfer, IN may efficiently catalyse other reactions: a third reaction, named disintegration, corresponds to the apparent inverse reaction of the strand transfer [2] although it is not clear whether it may occur in the cell context. More recently, a specific and internal cleavage catalysed by the full-length IN has been observed *in*

*vitro* [3]. This reaction requires a symmetrical organisation of the DNA substrate as well as a tetrameric organisation of the protein. From a structural point of view, this reaction is related to the endonucleolytic reaction of a restriction enzyme.

*In vivo*, the integrase oligomer and viral DNA molecule form part of a preintegration complex (PIC), our knowledge of which remains limited. The reverse transcriptase (RT), matrix protein (MA), Vpr and the nucleocapsid protein (NC) are also present in this complex as well as cellular partners [4-7]. The presence of an intact integrase is required for the stabilisation of preintegration complexes and their transport into the nucleus: These non catalytic functions of IN are also crucial for the viral replication cycle. Indeed, a functional interaction between IN and RT has been observed, suggesting that IN is involved, at least indirectly, in controlling the synthesis of viral DNA [8-10]. Furthermore, the interaction of particular IN structures with one or several cellular cofactors plays a key role for the integration into host cell chromosomes. For instance, LEDGF/p75 acts as a chromatin tethering factor for IN [11,12]. All these observations pave the way for the development of inhibitors targeting the interactions between IN and either viral or cellular cofactors. These alternative functions may constitute useful targets for the future development of integrase inhibitors.

## 1. Integrase

Integrase is a 288-amino acid protein (32 kDa) encoded by the end of the *pol* gene. It is produced as part of the Gag-Pol polypeptide precursor, from which it is released by viral protease-mediated cleavage. It has three independent domains: (i) The N-terminal domain (amino acids 1-49) that carries an HHCC motif analogous to a zinc finger, and effectively binds  $Zn^{2+}$  [13], possibly favouring protein multimerisation, a key process in integration [13,14]. (ii) The central domain or catalytic domain (amino acids 50-212) encompassing a D,D-35,E motif which is indispensable for the catalytic activity and which is conserved between viral IN and transposases. This central domain is also implicated in the binding of the viral DNA extremities mainly via the residues Q148, K156 and K159 [15-19]. All integrase activities strictly require the presence of a metallic cationic cofactor which is coordinated by two residues of the catalytic triad (D64 and D116 for HIV-1 IN) [20,21]. (iii) The C-terminal domain (amino acids 213-288) binds non-specifically to DNA and therefore is mainly involved in the stability of the complex with DNA. No complete structure has yet been determined for the integrase protomer ( $IN^{1-288}$ ), or for oligomers or complexes of these structures with DNA, due to poor solubility and interdomain flexibility problems. However, several structures of isolated domains or of two consecutive domains have been reported [20-25].

Integrase functions in a multimeric form, as shown by complementation experiments: mixtures of proteins, each individually inactive, were found to be active [26-28]. For example, an inactive catalytic triad mutant can be complemented by an inactive integrase truncated at its C-terminal end. Such a functional complementation

can be observed in virions [29]. In addition, the factors promoting integrase multimerisation such as  $Zn^{2+}$  also stimulate the specific  $Mg^{2+}$ -dependent activity of the enzyme [14], indicating that functional enzyme is multimeric. Dimers form at either end of the viral DNA molecule. These dimers are responsible for 3'-processing activity [30-34]. Pairs of dimers bring together the two ends of the viral DNA and leads to the formation of a tetramer (dimer of dimer), the active form for concerted integration [35,36]. During its catalytic cycle, IN must bind simultaneously to the viral substrate DNA and the target DNA. Current knowledge of the organisation of this tetramer onto DNA is based exclusively on models constructed from partial structural and biochemical (cross-linking and site-directed mutagenesis) data. In the most recent model, an IN tetramer is bound to the two ends of the viral DNA, *i.e.* LTRs (Long Terminal Repeat), and to a 26 base pairs host DNA molecule in the presence of  $Mg^{2+}$  [37]. This model takes into account the structural constraints deduced from the model of the complex formed between DNA and a related enzyme, the Tn5 transposase, and the observation that the two ends of the viral DNA are integrated five base pairs apart, corresponding to a distance of about 16 Å. This model may provide a platform for the rational design of new inhibitors.

## 2. The catalytic activities of integrase (IN)

### 2.1 *3'-processing and strand transfer*

There is now substantial virological evidence that the precursor of integrated viral DNA, or provirus, is a linear viral DNA generated by reverse transcription of the viral genome. Two reactions are required for the covalent integration of viral DNA into the

host DNA. The integrase (IN) first binds to a short sequence at each end of the viral DNA known as the long terminal repeat (LTR) and catalyses an endonucleotide cleavage known as 3'-processing, in which a dinucleotide is eliminated from each end of the viral DNA (Fig 1.). The resulting cleaved DNA is then used as a substrate for integration or strand transfer leading to the covalent insertion of the viral DNA into the genome of the infected cell (Fig 1.). This second reaction occurs simultaneously at both ends of the viral DNA molecule, with an offset of precisely five base pairs between the two opposite points of insertion.

*In vivo*, these two reactions are spatially and temporally separated. The ends of the viral DNA are processed in the cytoplasm, whereas strand transfer occurs in the nucleus. The two reactions are also energetically independent. In both cases, the reaction is a single-step *trans*-esterification involving the disruption of a phosphodiester bond by nucleophilic attack. In the first reaction, the bond concerned is part of the viral DNA molecule and in the second, the bond is in the target DNA. There is therefore no covalent intermediate between the enzyme and the DNA as it is observed during catalytic reaction of topoisomerase or IN of lambda phage, for example. The removal of the dinucleotides from the 5' overhang, of viral origin, and DNA repair (*i.e.* polymerisation and ligation) are required to complete the full integration reaction. One study suggested that this might involve a DNA-dependent DNA polymerase activity of the IN [38], but, to date, such a polymerase activity of IN was not confirmed and it is generally thought that this DNA repair is performed by cellular mechanisms that can be reproduced *in vitro* with purified host cell factors [39]. The final reaction thus results in a viral DNA molecule, the provirus, integrated into and collinear with the genomic DNA, with a characteristic 5 base pairs

duplication (in the case of HIV-1) of genomic sequence flanking the integration site. Several lines of evidence support a non-random integration with preferential integration in transcription units for HIV-1 [40]. Integration is then mainly directed by interactions between the pre-integration complex and chromatin (see # 3). From a DNA sequence point of view, it was recently shown that integration occurs preferentially within symmetric sequences [41-43] (see # 2.3).

Both reactions (3'-processing and strand transfer) can be reproduced *in vitro* using short double-stranded oligonucleotides mimicking the sequence of the ends of the viral LTR U5 or U3 in the presence of a recombinant integrase [44,45]. 3'-processing is a highly specific reaction. This reaction involves the removal of a dinucleotide, adjacent to the highly conserved CA dinucleotide, from the 3' strand of the U3 and U5 viral DNA LTRs. Mutations in this sequence completely abolish activity, whereas the integrity of flanking sequences is much less important [15,46]. The 3'-processing reaction corresponds to a nucleophilic attack by a water molecule. However, other alternative nucleophilic agents can be used such as glycerol but generally conduct to non-specific endonucleolytic cleavage [47-49]. This mainly occurs when  $Mn^{2+}$  is used. The 3'-OH of the unprocessed DNA can also be used directly as a nucleophilic agent leading to 3'-5' cyclic dinucleotide product [47]. The use of the physiological relevant cofactor  $Mg^{2+}$  improves the specificity of the cleavage with water as the mainly used nucleophilic agent.

Interestingly, during the same reaction, IN can catalyse, with a modest yield, the strand transfer. In the strand transfer reaction, the nucleophilic agent corresponds to the 3'-OH extremity of the processed strand. It is possible to increase the yield of

the strand transfer with pre-processed oligonucleotides [36]. By using an oligonucleotide mimicking one LTR end, only a half-transfer reaction can be observed. Concerted integration is more difficult to reproduce. *In vitro* experiments using long DNA fragments with two viral extremities have been performed demonstrating that IN alone was sufficient to catalyse the concerted reaction [35,36,50,51]. However, viral or cellular protein, acting as cofactors for the integration process, such as the viral nucleocapsid protein NC [52] and the cellular proteins HMG I(Y) [53] and LEDGF [54-56] may increase its efficacy. Interestingly, it was recently shown that, in contrast to the half-transfer reaction, a higher reaction yield was obtained for the concerted integration starting from a blunt-ended as compared to a pre-processed DNA substrate [36]. Furthermore, activity of IN is strongly dependent on its oligomeric state [14,44,57]. In contrast to 3'-processing which requires the dimeric form of IN [31], it was shown that concerted integration requires a tetrameric organization [32].

Both the 3'-processing and strand transfer reactions require a metallic cofactor. This cofactor may be  $Mn^{2+}$  or  $Mg^{2+}$ , but  $Mg^{2+}$  is preferentially used *in vivo*. Indeed, there is considerable experimental evidence to suggest that  $Mg^{2+}$  is more physiologically relevant, particularly as the specificity of the reaction is much greater in the presence of this cation: (i) IN displays strong non-specific nuclease activity in the presence of  $Mn^{2+}$  [58,59]. (ii) The tolerance of sequence variation at the ends of the viral DNA molecule is much greater in the presence of  $Mn^{2+}$  than in the presence of  $Mg^{2+}$  [15,46]. (iii) Many IN mutations remain silent in the presence of  $Mn^{2+}$  but not in the presence of  $Mg^{2+}$ . For example, mutations of the HHCC domain that are deleterious to the virus *in vivo* affect 3'-processing and integration activity in *in vitro*

tests using  $Mg^{2+}$ , but have no such effect in tests using  $Mn^{2+}$  [60,61]. Furthermore, zinc has no stimulatory effect on IN activity when using  $Mn^{2+}$  as a cofactor while zinc stimulates the  $Mg^{2+}$ -dependent activity [14]. The difference between cofactors has pharmacological implications, as the apparent efficacy of various IN inhibitors differs between tests using  $Mg^{2+}$  or  $Mn^{2+}$  as a cofactor [62-64], and the effects of mutations conferring drug resistance are often detectable only in tests using  $Mg^{2+}$  as the cofactor [65]. These considerations have led to the use of chemical groups chelating  $Mg^{2+}$  in the rational design of integrase inhibitors. Such groups are present in all the inhibitors developed to date, including raltegravir and elvitegravir [1].

Whatever the activity tested, IN is characterized by an overall slow cleavage efficiency. Furthermore, IN form stable complexes with both DNA substrate and DNA product, limiting multiple turnover [66]. Taken together, these features resemble to those observed for other polynucleotidyl-transferases such as transposases. These enzymes share a peculiar enzymatic property: they have evolved to catalyse multi-sequential steps (two reactions for IN and four for Tn5 transposase) in a single active site. A multi-sequential reaction requires a strong binding of the enzyme to the DNA product after each chemical step to optimise the entire process but consequently diminishes the overall enzymatic efficacy in term of turnover. However, this weak catalytic activity is not detrimental for these enzymes in the cellular context, because a single event of integration or transposition is sufficient for the overall function.

*In vivo*, the tight binding of IN to the viral DNA (indifferently unprocessed or processed) allows the complex to remain associated after the 3'-processing reaction long enough to be translocated into the nucleus for subsequent integration. Two

strategies have been considered for the development of IN inhibitors: screening using the unbound protein (before complex formation) or screening with the preformed IN-viral DNA complex. The success of these two approaches has been demonstrated by the identification of (i) inhibitors of 3'-processing targeting the DNA free enzyme and blocking its binding to the viral DNA [67] and (ii) inhibitors of strand transfer targeting the preformed complex more related to the preintegration complex (PIC) [68]. These two families of compounds are respectively called INBI (IN DNA-Binding Inhibitors) and INSTI (IN Strand Transfer Inhibitors) (Fig 2.). Since the early 1990s, a number of compounds inhibiting either 3'-processing or strand transfer have been identified *in vitro* [69,70]. The great stability of the PIC and its presence in the cell throughout most of the preintegration steps make this complex the most suitable target. Unfortunately, most of the INBI compounds are inactive on the preformed complexes. Indeed only strand transfer inhibitors or INSTIs have been shown to be potent antiviral compounds. As they selectively target the preformed IN-viral DNA complex and inhibit the binding of the acceptor DNA (*i.e.* target DNA or host DNA), INSTI compounds selectively inhibit the strand transfer reaction and have no effect on the 3'-processing reaction [71]. One such compound, Raltegravir (Isentress®), which was developed based on early studies by Hazuda et al. [68], was approved for clinical use in Autumn 2007 as the first antiretroviral drug (ARV) targeting the viral integrase (IN). This inhibitor act by binding to the IN-viral DNA complex, close to the 3' end of the donor DNA, thereby selectively blocking the strand transfer step; the IC<sub>50</sub> values are in the nanomolar range both *in vitro* and *ex vivo* with a high therapeutic index [72]. Unfortunately, variants of the virus resistant to this inhibitor have already been reported [72]. The emergence of resistant virus *in vivo* should

prompt both a search for new INSTIs and reassessment of the potential inhibitory activity of INBIs (such as styrylquinolines or SQL) which have been shown to be inhibitors of 3'-processing *in vitro* with significant inhibitory activity against viral replication in cell cultures (Fig 2.) [73].

The presence either of the catechol or an another group on the SQLs able to form a complex of coordination with a divalent ion suggests that these compounds interact with the active site of the enzyme by a chelation with the metallic cofactor. These compounds are mainly inhibitors of the 3'-processing reaction, and their mechanism of action *in vitro* can be assimilated to a competitive mechanism. Recently, experiments based on fluorescence anisotropy demonstrated that SQLs are DNA-binding inhibitors of HIV-1 IN [67]. In summary, INBI compounds primarily compete with the binding of the donor DNA (viral DNA) while INSTI compounds compete with the binding of the acceptor DNA (target DNA). However, the mechanism of inhibition of SQLs in the cell context is not completely understood. These compounds appear to act at steps prior to integration, more particularly during RT [74] and nuclear import [75]. These effects are mediated by IN as evidenced by the appearance of resistance mutation in IN sequence. It is then suggested that, *ex vivo*, non catalytic region of IN are targeted by SQs (see paragraph "non catalytic role of IN"). It is interesting to note that the two classes of IN inhibitors, INBI and INSTI, induce distinct resistant mutations [68,74,76-78] highlighting the necessity to improve both classes of compounds.

## 2.2. Disintegration

A third reaction, disintegration, is observed *in vitro* (Fig 1.). Disintegration may be considered to be the reverse of the strand transfer reaction [2]. Unlike the 3'-processing and strand transfer reactions which requires the full-length protein, the disintegration reaction can be catalysed by the catalytic domain alone (IN<sup>55-212</sup>) or by truncated proteins, IN<sup>1-212</sup> or IN<sup>55-288</sup> [44,79,80]. This activity was widely used for testing the competitive mechanisms of certain inhibitors. There is currently no experimental evidence to suggest that this reaction occurs *in vivo*.

## 2.3. A new internal specific activity

Recently, our group has identified a new internal and specific cleavage activity *in vitro* of HIV-1 IN [3]. Until now, all attempts to study a specific internal endonucleolytic cleavage *in vitro* have failed. Vink et al. have demonstrated that when the CA dinucleotide, indispensable for the 3'-processing, was separated by more than 2 nucleotides from the 3'-OH end, the activity was dramatically impaired [81]. Nevertheless, we have demonstrated that oligonucleotides mimicking the palindromic sequence found at the LTR-LTR junction of the 2-LTR circles (found in infected cells) were efficiently cleaved at internal positions by HIV-1 IN, with cleavage kinetics comparable to the 3'-processing reaction (Fig 1.). This reaction occurs symmetrically on both strands, with a strong cleavage at the CA dinucleotide (corresponding to the CA sequence used for the 3'-processing reaction). A second weaker cleavage site appears after the next adenine (TA sequence) in the 5'-3' direction. Furthermore, HIV-1 IN can efficiently cleave a plasmid mimicking the 2-LTR circles specifically at the LTR-LTR junction. The specificity of this reaction is similar to

the one catalysed by transposases which cleave the DNA substrate after a CA or TA dinucleotide [82]. Such internal cleavages are not observed using a mutant of the catalytic site (E152A) testifying that the DDE triad is also implicated in this reaction. In addition, this novel activity is stringent and highly specific as (i) it occurs with the physiological metallic cofactor ( $Mg^{2+}$ ) and not only  $Mn^{2+}$ , (ii) only the full-length IN is competent for the internal cleavage of the palindrome, in contrast to the disintegration reaction that is efficiently catalysed by truncated proteins such as IN<sup>55-212</sup>, IN<sup>55-288</sup>, IN<sup>1-212</sup> and (iii) it does not sustain any mutation in the sequence of the LTR-LTR junction. Furthermore, the cleavage of the LTR-LTR junction requires the tetrameric forms of IN whereas the 3'-processing reaction is efficiently catalysed by a dimer [31,32]. This new activity seems to be generalised to other retroviral IN as reported earlier for PFV-1 IN [83,84]. However, although PFV-1 IN performs this cleavage activity, it is important to note that both IN are strictly restricted to their own cognate palindromic sequence: HIV-1 IN is unable to cleave the PFV LTR-LTR junction and PFV-1 IN is unable to cleave the HIV-1 LTR-LTR junction.

Recently, mapping of extensive integration sites, notably for HIV-1, put in light the existence of a weak palindromic consensus [41-43]. It is important to note that the sequence of the weak palindromic consensus is similar, although not identical, to the one found at the LTR-LTR junction. This specific endonucleolytic activity on a palindromic LTR-LTR junction as well as the symmetrical organization of integration sites reveal a common structural feature of IN: IN intrinsically prefers to bind to symmetric DNA sequences. Moreover, we have found that tetramers catalyses the cleavage of the palindromic sequence while others have suggested that the same oligomeric form is responsible for the concerted integration in the context of the

synaptic complex [35,36]. Therefore, one could reasonably imagine that the same multimeric organization of IN (*i.e.* the tetrameric form) is stabilised by a corresponding symmetry at the DNA level, either at the viral DNA (LTR-LTR junction) or at the target level (integration sites).

*In vivo*, unintegrated viral DNA could represent 99% of total viral DNA in infected cells [85] underlying that integration is a rare event. Un-integrated DNA is mainly linear but also circular - 1-LTR or 2-LTR circles. In the absence of integration (for example using strand-transfer inhibitors such as diketo acids), at least the 2-LTR circular forms of viral DNA, which are usually believed to be dead-end molecules, are accumulated [86]. It is tempting to speculate about a possible role of 2-LTR circles in a subsequent integration process after removing the drug pressure, mediated by the ability of IN to cleave the LTR-LTR junction. However, to date, although IN is able to cleave the LTR-LTR junction *in vitro*, there is no proof that such a cleavage can occur *in vivo* and thus that 2-LTR circles could be an efficient precursor for integration.

### 3. Non catalytic activities of IN

IN plays a key role for retroviral replication because of its catalytical activities. In addition, IN has also non catalytic properties that are essential for the replication cycle. Mutations of IN can be divided into two groups. The first class of mutations (Class I) includes mutants that are affected in their catalytic activities. For instance, one mutation in either the three amino acids of the DDE triad abolishes the catalytic activities of IN. The second class of mutations (Class II mutants) disturbs other steps of the retroviral replication and corresponding purified integrase mutants display wild-type level of activity.

Several mutations of IN displayed an *in vivo* DNA synthesis defect and a block of viral replication at the reverse transcription level [8,9,87-89]. A structural general defect at the level of the retrotranscription complex which includes RT and IN may account for such a phenotype. Indeed, several studies suggest a physical interaction between IN and RT [9]. Such a defect in DNA synthesis can be also observed when using SQL compounds which target integrase as evidenced by resistance mutations study but primarily affect the reverse transcription step [74].

Another role of IN prior to integration is related to the PIC translocation in the nucleus. In fact, in non-dividing infected cells, such as macrophages, the PIC must cross the nuclear membrane to reach the chromosomal DNA. This involves an active mechanism, the determinants of which remain unclear [90,91]. Integrase and at least two other components of the PIC, Vpr and MA, have karyophilic properties [92]. The integrase enzyme includes several sequence motifs likely to act as nuclear localisation signals (NLSs), including at least one known to interact with the nuclear import receptor, this motif being located in the C-terminal domain [91]. A sequence within the catalytic core including the V165 and R166 residues may also contribute to the karyophilic properties of integrase [93], although this remains a matter of debate [89,94]. In any case, the mutation of these various sequences does not completely abolish the nuclear translocation of PICs, suggesting that there are complementary and/or redundant translocation mechanisms.

Additionally, interactions between integrase and cellular protein partners play key role in the targeting of integration. A systematic study of the sites of HIV DNA integration into the host DNA has shown that integration is not entirely random.

Analysis of integration sites *in vivo* indicates that HIV tends to integrate into sites of active transcription [40]. It is likely that this integration bias results from interactions between PICs and components of cellular origin in relationship with the chromatin tethering. Several cellular factors, including Ini-1 [95,96], BAF [97,98], Ku [99] and LEDGF/p75 [100], are known to interact with the PIC in the nucleus. Among these proteins, at least Ini-1 and LEDGF/p75 physically interact with IN [95,101]. Recent work with LEDGF/p75 strongly suggests that this factor is effectively responsible for targeting integration [11]. LEDGF/p75 silencing modifies the bias from transcription units to CpG islands [40,102]. Given that LEDGF/p75 is essential for HIV-1 replication and LEDGF/p75 interacts directly with integrase, the domain of interaction between these two proteins is therefore a promising target for the development of integrase ligands with antiviral activity. Ini-1, which has homology with the yeast SNF5 component of the SWI/SNF complex involved in the remodelling of the chromatin, may also be involved in this targeting of integration [103]. Although no direct interaction between IN and BAF or Ku was described, it is suggested that these two factors could influence the profile or efficiency of integration [104,105]. Access to chromatin may be favoured by the presence in the PIC of the cellular factor BAF (barrier to autointegration factor) and the interaction of this factor with another cellular factor, emerin, which binds the internal face of the nuclear membrane to chromatin [106]. In relationship with chromatin, it was recently described that the C-terminal domain of integrase is acetylated by a histone acetyl transferase (HAT) [107]. However, the effect of IN acetylation on integration *in vivo* remains unclear [108].

Finally, integrase could be indirectly involved in the regulation of transcription of integrated provirus. After the integration process, IN could be tightly bound to the

integrated DNA and then, the degradation of IN by the proteasome-ubiquitin pathway was proposed to regulate the transcription of viral genes. Indeed, Dargemont and collaborators have found that integrase interacts with VBP1 (von Hippel-Lindau binding protein 1), a binding partner of Cul2/VHL ligase complex involved in the polyubiquitylation process [109].

## CONCLUSION

In conclusion, remarkable progress has been made towards understanding the structure of the preintegration complex formed by HIV integrase and viral DNA. This new knowledge has led to considerable improvements in the methods used to search for compounds active against this enzyme. Several families of inhibitors have now been identified, including at least one — strand transfer inhibitors — currently in the advanced stages of clinical development and giving results sufficiently promising for one molecule (raltegravir) to have obtained a licence in October 2007 for release in the United States. The identification of several integrase cofactors has provided us with a clearer picture of the determinants of integration *in vivo*, opening up new possibilities for pharmacological research. There is no doubt that interest in the structural biology of integrase will be substantially stimulated by the release of the first integrase inhibitors onto the market and, unfortunately, by the likely emergence of resistant viruses.

Abbreviations:

HIV-1: Human Immunodeficiency virus type 1

PFV-1: Primate Foamy virus type 1

MK-0518: Raltegravir

ARVs: Antiretroviral drugs

IN: Integrase

RT: Reverse Transcriptase

MA: Matrix

NC: Nucleocapsid

LTR: Long Terminal Repeat

INBI: IN DNA-Binding Inhibitor

INSTI: IN Strand Transfer Inhibitor

PIC: Preintegration complex

LEDGF: Lens Epithelial Derived Growth Factor

Competing interests:

The authors declare that they have no competing interests.

## Legend Figures

Figure 1. Catalytical activities of HIV-1 integrase. The catalytical activities 3'-processing (A), strand-transfer reaction (B), disintegration (C) and palindrome cleavage (D) are represented. The domains of the protein responsible for these activities are depicted in the table above.

Figure 2. Some anti-integrase compounds. Styrylquinoline, a member of the INBI (IN DNA-Binding Inhibitors) compound and beta dicetonc acid, Raltegravir and Elvitegravir, members of the INSTI (IN Strand Transfer Inhibitors) compounds, are represented.

## Reference List

1. Al Mawsawi LQ, Al Safi RI, Neamati N: Anti-infectives clinical progress of HIV-1 integrase inhibitors. *Expert Opin Emerg Drugs* 2008, 13: 213-225.
2. Chow SA, Vincent KA, Ellison V, Brown PO: Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* 1992, 255: 723-726.
3. Delelis O, Parissi V, Leh H, Mbemba G, Petit C, Sonigo P, Deprez E, Mouscadet JF: Efficient and specific internal cleavage of a retroviral palindromic DNA sequence by tetrameric HIV-1 integrase. *PLoS ONE* 2007, 2: e608.
4. Miller MD, Farnet CM, Bushman FD: Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* 1997, 71: 5382-5390.
5. Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M: 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* 1993, 90: 6125-6129.
6. Nermut MV, Fassati A: Structural analyses of purified human immunodeficiency virus type 1 intracellular reverse transcription complexes. *J Virol* 2003, 77: 8196-8206.

7. Gallay P, Swingler S, Song J, Bushman F, Trono D: HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell* 1995, 83: 569-576.
8. Wu X, Liu H, Xiao H, Conway JA, Hehl E, Kalpana GV *et al.*: Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions with the nucleoprotein reverse transcription complex. *J Virol* 1999, 73: 2126-2135.
9. Zhu K, Dobard C, Chow SA: Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase. *J Virol* 2004, 78: 5045-5055.
10. Dobard CW, Briones MS, Chow SA: Molecular mechanisms by which human immunodeficiency virus type 1 integrase stimulates the early steps of reverse transcription. *J Virol* 2007, 81: 10037-10046.
11. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH *et al.*: An Essential Role for LEDGF/p75 in HIV Integration. *Science* 2006.
12. Hombrouck A, De Rijck J, Hendrix J, Vandekerckhove L, Voet A, De Maeyer M *et al.*: Virus evolution reveals an exclusive role for LEDGF/p75 in chromosomal tethering of HIV. *PLoS Pathog* 2007, 3: e47.
13. Zheng R, Jenkins TM, Craigie R: Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci U S A* 1996, 93: 13659-13664.
14. Lee SP, Xiao J, Knutson JR, Lewis MS, Han MK: Zn<sup>2+</sup> promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. *Biochemistry* 1997, 36: 173-180.
15. Esposito D, Craigie R: Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J* 1998, 17: 5832-5843.
16. Jenkins TM, Esposito D, Engelman A, Craigie R: Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J* 1997, 16: 6849-6859.
17. Heuer TS, Brown PO: 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* 1997, 36: 10655-10665.
18. Drake RR, Neamati N, Hong H, Pilon AA, Sunthankar P, Hume SD *et al.*: Identification of a nucleotide binding site in HIV-1 integrase. *Proc Natl Acad Sci U S A* 1998, 95: 4170-4175.
19. Johnson AA, Santos W, Pais GC, Marchand C, Amin R, Burke TR, Jr. *et al.*: Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J Biol Chem* 2006, 281: 461-467.
20. Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR: Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc Natl Acad Sci U S A* 1998, 95: 9150-9154.

21. Maignan S, Guilloteau JP, Zhou-Liu Q, Clement-Mella C, Mikol V: Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J Mol Biol* 1998, 282: 359-368.
22. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM: Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat Struct Biol* 1997, 4: 567-577.
23. Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R *et al.*: Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* 1995, 34: 9826-9833.
24. Wang JY, Ling H, Yang W, Craigie R: Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J* 2001, 20: 7333-7343.
25. Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, Leavitt AD *et al.*: Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc Natl Acad Sci U S A* 2000, 97: 8233-8238.
26. van Gent DC, Vink C, Groeneger AA, Plasterk RH: Complementation between HIV integrase proteins mutated in different domains. *EMBO J* 1993, 12: 3261-3267.
27. Engelman A, Bushman FD, Craigie R: Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J* 1993, 12: 3269-3275.
28. van den Ent FM, Vos A, Plasterk RH: Dissecting the role of the N-terminal domain of human immunodeficiency virus integrase by trans-complementation analysis. *J Virol* 1999, 73: 3176-3183.
29. Fletcher TM, III, Soares MA, McPhearson S, Hui H, Wiskerchen M, Muesing MA *et al.*: Complementation of integrase function in HIV-1 virions. *EMBO J* 1997, 16: 5123-5138.
30. Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, Hawkins ME *et al.*: DNA binding induces dissociation of the multimeric form of HIV-1 integrase: a time-resolved fluorescence anisotropy study. *Proc Natl Acad Sci U S A* 2001, 98: 10090-10095.
31. Guiot E, Carayon K, Delelis O, Simon F, Tauc P, Zubin E, Gottikh M, Mouscadet JF, Brochon JC, Deprez E: Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. *J Biol Chem* 2006, 281: 22707-22719.
32. Faure A, Calmels C, Desjobert C, Castroviejo M, Caumont-Sarcos A, Tarrago-Litvak L *et al.*: HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res* 2005, 33: 977-986.
33. Baranova S, Tuzikov FV, Zakharova OD, Tuzikova NA, Calmels C, Litvak S *et al.*: Small-angle X-ray characterization of the nucleoprotein complexes resulting from DNA-induced oligomerization of HIV-1 integrase. *Nucleic Acids Res* 2007, 35: 975-987.
34. Delelis O, Carayon K, Guiot E, Leh H, Tauc P, Brochon JC, Mouscadet JF, Deprez E: Insight into the integrase-DNA recognition mechanism: A specific DNA-binding mode revealed by an enzymatically labeled integrase. *J Biol Chem* 2008.

35. Li M, Mizuuchi M, Burke TR, Jr., Craigie R: Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J* 2006, 25: 1295-1304.
36. Li M, Craigie R: Processing of viral DNA ends channels the HIV-1 integration reaction to concerted integration. *J Biol Chem* 2005, 280: 29334-29339.
37. Wielens J, Crosby IT, Chalmers DK: A three-dimensional model of the human immunodeficiency virus type 1 integration complex. *J Comput Aided Mol Des* 2005, 19: 301-317.
38. Acel A, Udashkin BE, Wainberg MA, Faust EA: Efficient gap repair catalyzed in vitro by an intrinsic DNA polymerase activity of human immunodeficiency virus type 1 integrase. *J Virol* 1998, 72: 2062-2071.
39. Brin E, Yi J, Skalka AM, Leis J: Modeling the late steps in HIV-1 retroviral integrase-catalyzed DNA integration. *J Biol Chem* 2000, 275: 39287-39295.
40. Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D *et al.*: Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS ONE* 2007, 2: e1340.
41. Grandgenett DP: Symmetrical recognition of cellular DNA target sequences during retroviral integration. *Proc Natl Acad Sci U S A* 2005, 102: 5903-5904.
42. Holman AG, Coffin JM: Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc Natl Acad Sci U S A* 2005, 102: 6103-6107.
43. Wu X, Li Y, Crise B, Burgess SM, Munroe DJ: Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J Virol* 2005, 79: 5211-5214.
44. Leh H, Brodin P, Bischerour J, Deprez E, Tauc P, Brochon JC *et al.*: Determinants of Mg<sup>2+</sup>-dependent activities of recombinant human immunodeficiency virus type 1 integrase. *Biochemistry* 2000, 39: 9285-9294.
45. Pinskaya M, Romanova E, Volkov E, Deprez E, Leh H, Brochon JC *et al.*: HIV-1 integrase complexes with DNA dissociate in the presence of short oligonucleotides conjugated to acridine. *Biochemistry* 2004, 43: 8735-8743.
46. Agapkina J, Smolov M, Barbe S, Zubin E, Zatsepin T, Deprez E *et al.*: Probing of HIV-1 integrase/DNA interactions using novel analogs of viral DNA. *J Biol Chem* 2006, 281: 11530-11540.
47. Engelman A, Mizuuchi K, Craigie R: HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 1991, 67: 1211-1221.
48. Gerton JL, Herschlag D, Brown PO: Stereospecificity of reactions catalyzed by HIV-1 integrase. *J Biol Chem* 1999, 274: 33480-33487.
49. Skinner LM, Sudol M, Harper AL, Katzman M: Nucleophile selection for the endonuclease activities of human, ovine, and avian retroviral integrases. *J Biol Chem* 2001, 276: 114-124.

50. Sinha S, Pursley MH, Grandgenett DP: Efficient concerted integration by recombinant human immunodeficiency virus type 1 integrase without cellular or viral cofactors. *J Virol* 2002, 76: 3105-3113.
51. Sinha S, Grandgenett DP: Recombinant human immunodeficiency virus type 1 integrase exhibits a capacity for full-site integration in vitro that is comparable to that of purified preintegration complexes from virus-infected cells. *J Virol* 2005, 79: 8208-8216.
52. Carteau S, Gorelick RJ, Bushman FD: Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. *J Virol* 1999, 73: 6670-6679.
53. Hindmarsh P, Ridky T, Reeves R, Andrade M, Skalka AM, Leis J: HMG protein family members stimulate human immunodeficiency virus type 1 and avian sarcoma virus concerted DNA integration in vitro. *J Virol* 1999, 73: 2994-3003.
54. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y *et al.*: HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* 2003, 278: 372-381.
55. Cherepanov P: LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res* 2007, 35: 113-124.
56. Pandey KK, Sinha S, Grandgenett DP: Transcriptional co-activator LEDGF/p75 modulates HIV-1 integrase mediated concerted integration. *J Virol* 2007.
57. Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, Brochon JC: Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. *Biochemistry* 2000, 39: 9275-9284.
58. Engelman A, Craigie R: Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity. *J Virol* 1995, 69: 5908-5911.
59. Bushman FD, Wang B: Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding. *J Virol* 1994, 68: 2215-2223.
60. Wiskerchen M, Muesing MA: Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. *J Virol* 1995, 69: 376-386.
61. Khan E, Mack JP, Katz RA, Kulkosky J, Skalka AM: Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res* 1991, 19: 851-860.
62. Fesen MR, Pommier Y, Leteurtre F, Hiroguchi S, Yung J, Kohn KW: Inhibition of HIV-1 integrase by flavones, caffeic acid phenethyl ester (CAPE) and related compounds. *Biochem Pharmacol* 1994, 48: 595-608.
63. Mazumder A, Neamati N, Ojwang JO, Sunder S, Rando RF, Pommier Y: Inhibition of the human immunodeficiency virus type 1 integrase by guanosine quartet structures. *Biochemistry* 1996, 35: 13762-13771.

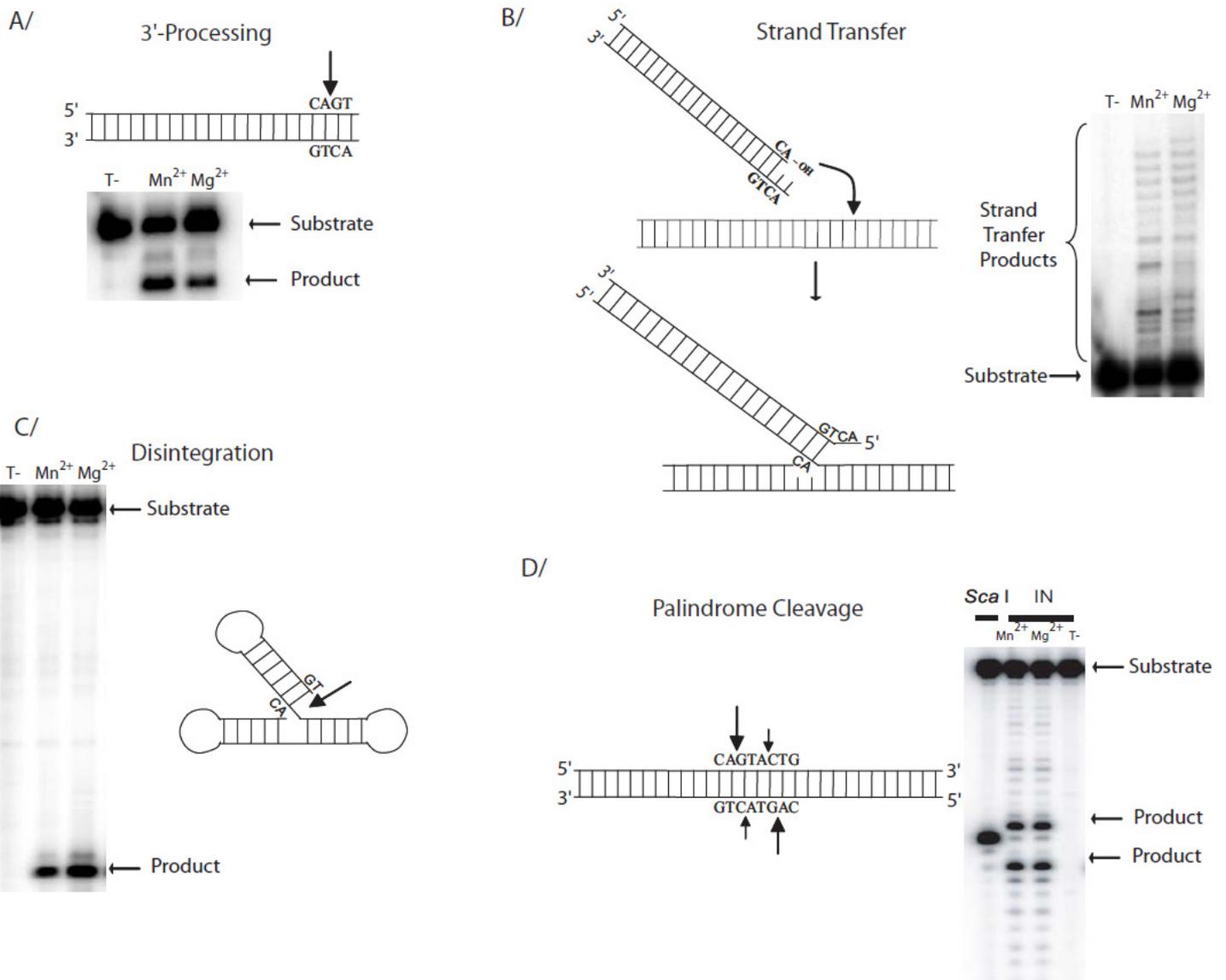
64. Molteni V, Rhodes D, Rubins K, Hansen M, Bushman FD, Siegel JS: A new class of HIV-1 integrase inhibitors: the 3,3,3', 3'-tetramethyl-1,1'-spirobi(indan)-5,5',6,6'-tetrol family. *J Med Chem* 2000, 43: 2031-2039.
65. Grobler JA, Stillmock K, Hu B, Witmer M, Felock P, Espeseth AS *et al.*: Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. *Proc Natl Acad Sci U S A* 2002, 99: 6661-6666.
66. Smolov M, Gottikh M, Tashlitskii V, Korolev S, Demidyuk I, Brochon JC, Mouscadet JF, Deprez E: Kinetic study of the HIV-1 DNA 3'-end processing. *FEBS J* 2006, 273: 1137-1151.
67. Deprez E, Barbe S, Kolaski M, Leh H, Zouhiri F, Auclair C *et al.*: Mechanism of HIV-1 integrase inhibition by styrylquinoline derivatives in vitro. *Mol Pharmacol* 2004, 65: 85-98.
68. Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler JA *et al.*: Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* 2000, 287: 646-650.
69. Semenova EA, Marchand C, Pommier Y: HIV-1 integrase inhibitors: update and perspectives. *Adv Pharmacol* 2008, 56: 199-228.
70. Egbertson MS: HIV integrase inhibitors: from diketoacids to heterocyclic templates: a history of HIV integrase medicinal chemistry at Merck West Point and Merck Rome (IRBM). *Curr Top Med Chem* 2007, 7: 1251-1272.
71. Espeseth AS, Felock P, Wolfe A, Witmer M, Grobler J, Anthony N *et al.*: HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc Natl Acad Sci U S A* 2000, 97: 11244-11249.
72. Malet I, Delelis O, Valantin MA, Montes B, Soulie C, Wirden M *et al.*: Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. *Antimicrob Agents Chemother* 2008, 52: 1351-1358.
73. Zouhiri F, Mouscadet JF, Mekouar K, Desmaele D, Savoure D, Leh H *et al.*: Structure-activity relationships and binding mode of styrylquinolines as potent inhibitors of HIV-1 integrase and replication of HIV-1 in cell culture. *J Med Chem* 2000, 43: 1533-1540.
74. Bonnenfant S, Thomas CM, Vita C, Subra F, Deprez E, Zouhiri F *et al.*: Styrylquinolines, integrase inhibitors acting prior to integration: a new mechanism of action for anti-integrase agents. *J Virol* 2004, 78: 5728-5736.
75. Mousnier A, Leh H, Mouscadet JF, Dargemont C: Nuclear import of HIV-1 integrase is inhibited in vitro by styrylquinoline derivatives. *Mol Pharmacol* 2004, 66: 783-788.
76. Hazuda DJ, Young SD, Guare JP, Anthony NJ, Gomez RP, Wai JS *et al.*: Integrase inhibitors and cellular immunity suppress retroviral replication in rhesus macaques. *Science* 2004, 305: 528-532.
77. Hazuda DJ, Anthony NJ, Gomez RP, Jolly SM, Wai JS, Zhuang L *et al.*: A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc Natl Acad Sci U S A* 2004, 101: 11233-11238.

78. Fikkert V, Van Maele B, Vercammen J, Hantson A, Van Remoortel B, Michiels M *et al.*: Development of resistance against diketo derivatives of human immunodeficiency virus type 1 by progressive accumulation of integrase mutations. *J Virol* 2003, 77: 11459-11470.
79. Gerton JL, Brown PO: The core domain of HIV-1 integrase recognizes key features of its DNA substrates. *J Biol Chem* 1997, 272: 25809-25815.
80. Laboulais C, Deprez E, Leh H, Mouscadet JF, Brochon JC, Le Bret M: HIV-1 integrase catalytic core: molecular dynamics and simulated fluorescence decays. *Biophys J* 2001, 81: 473-489.
81. Vink C, van Gent DC, Elgersma Y, Plasterk RH: Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. *J Virol* 1991, 65: 4636-4644.
82. Lee I, Harshey RM: Importance of the conserved CA dinucleotide at Mu termini. *J Mol Biol* 2001, 314: 433-444.
83. Delelis O, Petit C, Leh H, Mbemba G, Mouscadet JF, Sonigo P: A novel function for spumaretrovirus integrase: an early requirement for integrase-mediated cleavage of 2 LTR circles. *Retrovirology* 2005, 2: 31.
84. Delelis O, Lehmann-Che J, Saib A: Foamy viruses--a world apart. *Curr Opin Microbiol* 2004, 7: 400-406.
85. Chun TW, Carruth L, Finzi D, Shen X, DiGiuseppe JA, Taylor H *et al.*: Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* 1997, 387: 183-188.
86. Svarovskaia ES, Barr R, Zhang X, Pais GC, Marchand C, Pommier Y *et al.*: Azido-containing diketo acid derivatives inhibit human immunodeficiency virus type 1 integrase in vivo and influence the frequency of deletions at two-long-terminal-repeat-circle junctions. *J Virol* 2004, 78: 3210-3222.
87. Leavitt AD, Robles G, Alesandro N, Varmus HE: Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection. *J Virol* 1996, 70: 721-728.
88. Tsurutani N, Kubo M, Maeda Y, Ohashi T, Yamamoto N, Kannagi M *et al.*: Identification of critical amino acid residues in human immunodeficiency virus type 1 IN required for efficient proviral DNA formation at steps prior to integration in dividing and nondividing cells. *J Virol* 2000, 74: 4795-4806.
89. Lu R, Limon A, Devroe E, Silver PA, Cherepanov P, Engelman A: Class II integrase mutants with changes in putative nuclear localization signals are primarily blocked at a postnuclear entry step of human immunodeficiency virus type 1 replication. *J Virol* 2004, 78: 12735-12746.
90. Depienne C, Mousnier A, Leh H, Le Rouzic E, Dormont D, Benichou S *et al.*: Characterization of the nuclear import pathway for HIV-1 integrase. *J Biol Chem* 2001, 276: 18102-18107.

91. Ao Z, Huang G, Yao H, Xu Z, Labine M, Cochrane AW *et al.*: Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. *J Biol Chem* 2007, 282: 13456-13467.
92. Bukrinsky M: A hard way to the nucleus. *Mol Med* 2004, 10: 1-5.
93. Bouyac-Bertoia M, Dvorin JD, Fouchier RA, Jenkins Y, Meyer BE, Wu LI *et al.*: HIV-1 infection requires a functional integrase NLS. *Mol Cell* 2001, 7: 1025-1035.
94. Dvorin JD, Bell P, Maul GG, Yamashita M, Emerman M, Malim MH: Reassessment of the roles of integrase and the central DNA flap in human immunodeficiency virus type 1 nuclear import. *J Virol* 2002, 76: 12087-12096.
95. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP: Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* 1994, 266: 2002-2006.
96. Yung E, Sorin M, Wang EJ, Perumal S, Ott D, Kalpana GV: Specificity of interaction of IN1/hSNF5 with retroviral integrases and its functional significance. *J Virol* 2004, 78: 2222-2231.
97. Zheng R, Ghirlando R, Lee MS, Mizuuchi K, Krause M, Craigie R: Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex. *Proc Natl Acad Sci U S A* 2000, 97: 8997-9002.
98. Lin CW, Engelman A: The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. *J Virol* 2003, 77: 5030-5036.
99. Lewinski MK, Bushman FD: Retroviral DNA integration--mechanism and consequences. *Adv Genet* 2005, 55: 147-181.
100. Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, Peretz M *et al.*: LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J Virol* 2004, 78: 9524-9537.
101. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A: Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc Natl Acad Sci U S A* 2005, 102: 17308-17313.
102. Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, Kellam P *et al.*: LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev* 2007, 21: 1767-1778.
103. Mahmoudi T, Parra M, Vries RG, Kauder SE, Verrijzer CP, Ott M *et al.*: The SWI/SNF chromatin-remodeling complex is a cofactor for Tat transactivation of the HIV promoter. *J Biol Chem* 2006, 281: 19960-19968.
104. Lee MS, Craigie R: A previously unidentified host protein protects retroviral DNA from autointegration. *Proc Natl Acad Sci U S A* 1998, 95: 1528-1533.

- 
105. Masson C, Bury-Mone S, Guiot E, Saez-Cirion A, Schoevaert-Brossault D, Brachet-Ducos C *et al.*: Ku80 participates in the targeting of retroviral transgenes to the chromatin of CHO cells. *J Virol* 2007, 81: 7924-7932.
  106. Jacque JM, Stevenson M: The inner-nuclear-envelope protein emerlin regulates HIV-1 infectivity. *Nature* 2006, 441: 641-645.
  107. Cereseto A, Manganaro L, Gutierrez MI, Terreni M, Fittipaldi A, Lusic M *et al.*: Acetylation of HIV-1 integrase by p300 regulates viral integration. *EMBO J* 2005, 24: 3070-3081.
  108. Topper M, Luo Y, Zhadina M, Mohammed K, Smith L, Muesing MA: Posttranslational acetylation of the human immunodeficiency virus type 1 integrase carboxyl-terminal domain is dispensable for viral replication. *J Virol* 2007, 81: 3012-3017.
  109. Mousnier A, Kubat N, Massias-Simon A, Segeral E, Rain JC, Benarous R *et al.*: von Hippel Lindau binding protein 1-mediated degradation of integrase affects HIV-1 gene expression at a postintegration step. *Proc Natl Acad Sci U S A* 2007, 104: 13615-13620.

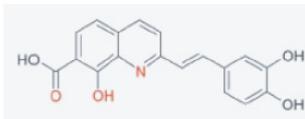
Fig 1.



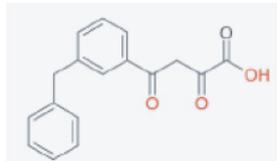
	Disintegration		3'-Processing		Strand Transfer		Palindrome Cleavage	
	Mn <sup>2+</sup>	Mg <sup>2+</sup>	Mn <sup>2+</sup>	Mg <sup>2+</sup>	Mn <sup>2+</sup>	Mg <sup>2+</sup>	Mn <sup>2+</sup>	Mg <sup>2+</sup>
N-Ter+CC	+	-	-	-	-	-	-	-
C-Ter+CC	+	-	-	-	-	-	-	-
CC	+	-	-	-	-	-	-	-
Full length	+	+	+	+	+	+	+	+

Fig 2.

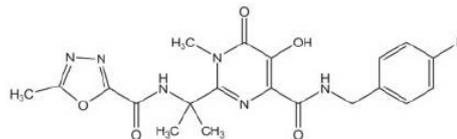
Styrylquinoline



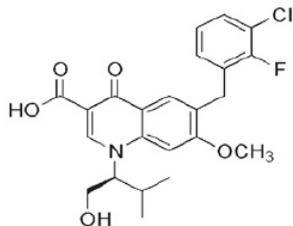
Beta-dicetonic acid



Raltegravir (MK-0518)



Elvitegravir (JTK-303)



# **BIBLIOGRAPHIE**

1. Acel A, Udashkin BE, Wainberg MA, and Faust EA (1998) Efficient gap repair catalyzed in vitro by an intrinsic DNA polymerase activity of human immunodeficiency virus type 1 integrase. *J Virol*, **72**, 2062-2071.
2. Ao Z, Fowke KR, Cohen EA, and Yao X (2005) 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.
3. Arhel N, Genovesio A, Kim KA, Miko S, Perret E, Olivo-Marin JC, Shorte S, and Charneau P (2006) Quantitative four-dimensional tracking of cytoplasmic and nuclear HIV-1 complexes. *Nat Methods*, **3**, 817-824.
4. Asante-Appiah E, Seeholzer SH, and Skalka AM (1998) Structural determinants of metal-induced conformational changes in HIV-1 integrase. *J Biol Chem*, **273**, 35078-35087.
5. Asante-Appiah E and Skalka AM (1997) A metal-induced conformational change and activation of HIV-1 integrase. *J Biol Chem*, **272**, 16196-16205.
6. Baekelandt V, Claeys A, Cherepanov P, De Clercq E, De Strooper B, Nuttin B, and Debysers Z (2000) DNA-Dependent protein kinase is not required for efficient lentivirus integration. *J Virol*, **74**, 11278-11285.
7. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dautet C, Xer-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, and Montagnier L (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868-871.
8. Basu VP, Song M, Gao L, Rigby ST, Hanson MN, and Bambara RA (2008) Strand transfer events during HIV-1 reverse transcription. *Virus Res*, **134**, 19-38.
9. Bebenek K, Abbotts J, Wilson SH, and Kunkel TA (1993) Error-prone polymerization by HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. *J Biol Chem*, **268**, 10324-10334.
10. Beese LS and Steitz TA (1991) Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J*, **10**, 25-33.
11. Bernardi F, Bottoni A, De Vivo M, Garavelli M, Keseru G, and Naray-Szabo G. A hypothetical mechanism for HIV-1 integrase catalytic action: DFT modelling of a bio-mimetic environment. *Chemical Physics Letters* **362**, 1-7. 2002.  
Ref Type: Generic
12. Berthoux L, Pechoux C, and Darlix JL (1999) Multiple effects of an anti-human immunodeficiency virus nucleocapsid inhibitor on virus morphology and replication. *J Virol*, **73**, 10000-10009.
13. Bhasin A, Goryshin IY, and Reznikoff WS (1999) Hairpin formation in Tn5 transposition. *J Biol Chem*, **274**, 37021-37029.

14. Bischerour J, Leh H, Deprez E, Brochon JC, and Mouscadet JF (2003) Disulfide-linked integrase oligomers involving C280 residues are formed in vitro and in vivo but are not essential for human immunodeficiency virus replication. *J Virol*, **77**, 135-141.
15. Bonnenfant S, Thomas CM, Vita C, Subra F, Deprez E, Zouhiri F, Desmaele D, d'Angelo J, Mouscadet JF, and Leh H (2004) Styrylquinolines, integrase inhibitors acting prior to integration: a new mechanism of action for anti-integrase agents. *J Virol*, **78**, 5728-5736.
16. Bosserman MA, O'Quinn DF, and Wong I (2007) Loop202-208 in avian sarcoma virus integrase mediates tetramer assembly and processing activity. *Biochemistry*, **46**, 11231-11239.
17. Bouyac-Bertoia M, Dvorin JD, Fouchier RA, Jenkins Y, Meyer BE, Wu LI, Emerman M, and Malim MH (2001) HIV-1 infection requires a functional integrase NLS. *Mol Cell*, **7**, 1025-1035.
18. Brochon JC (1994) Maximum entropy method of data analysis in time-resolved spectroscopy. *Methods Enzymol*, **240**, 262-311.
19. Brussel A and Sonigo P (2004) Evidence for gene expression by unintegrated human immunodeficiency virus type 1 DNA species. *J Virol*, **78**, 11263-11271.
20. Bujacz G, Alexandratos J, Wlodawer A, Merkel G, Andrade M, Katz RA, and Skalka AM (1997) Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J Biol Chem*, **272**, 18161-18168.
21. Bukrinsky M (2004) A hard way to the nucleus. *Mol Med*, **10**, 1-5.
22. Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, and Stevenson M (1993) 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.
23. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, and Gronenborn AM (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat Struct Biol*, **4**, 567-577.
24. Cannon PM, Byles ED, Kingsman SM, and Kingsman AJ (1996) Conserved sequences in the carboxyl terminus of integrase that are essential for human immunodeficiency virus type 1 replication. *J Virol*, **70**, 651-657.
25. Cannon PM, Wilson W, Byles E, Kingsman SM, and Kingsman AJ (1994) Human immunodeficiency virus type 1 integrase: effect on viral replication of mutations at highly conserved residues. *J Virol*, **68**, 4768-4775.
26. Carteau S, Gorelick RJ, and Bushman FD (1999) Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. *J Virol*, **73**, 6670-6679.

27. Cavert W and Haase AT (1998) A national tissue bank to track HIV eradication and immune reconstruction. *Science*, **280**, 1865-1866.
28. Cereseto A, Manganaro L, Gutierrez MI, Terreni M, Fittipaldi A, Lusic M, Marcello A, and Giacca M (2005) Acetylation of HIV-1 integrase by p300 regulates viral integration. *EMBO J*, **24**, 3070-3081.
29. Chan DC and Kim PS (1998) HIV entry and its inhibition. *Cell*, **93**, 681-684.
30. Charles W.Bock, Amy Kaufman Katz, George D.Markham, and Jenny P.Glusker. **Manganese as a Replacement for Magnesium and Zinc: Functional Comparison of the Divalent Ions** . *J.Am.Chem.Soc.* **121** (32), 7360-7372. 1999.  
Ref Type: Generic
31. Chen A, Weber IT, Harrison RW, and Leis J (2006) 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.
32. Chen H and Engelman A (1998) The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc Natl Acad Sci U S A*, **95**, 15270-15274.
33. Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, Leavitt AD, and Stroud RM (2000a) 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.
34. Chen X, Tsiang M, Yu F, Hung M, Jones GS, Zeynalzadegan A, Qi X, Jin H, Kim CU, Swaminathan S, and Chen JM (2008) Modeling, analysis, and validation of a novel HIV integrase structure provide insights into the binding modes of potent integrase inhibitors. *J Mol Biol*, **380**, 504-519.
35. Chen Z, Yan Y, Munshi S, Li Y, Zugay-Murphy J, Xu B, Witmer M, Felock P, Wolfe A, Sardana V, Emini EA, Hazuda D, and Kuo LC (2000b) X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50-293)--an initial glance of the viral DNA binding platform. *J Mol Biol*, **296**, 521-533.
36. Cherepanov P (2007) LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res*, **35**, 113-124.
37. Cherepanov P, Este JA, Rando RF, Ojwang JO, Reekmans G, Steinfeld R, David G, De Clercq E, and Debyser Z (1997) Mode of interaction of G-quartets with the integrase of human immunodeficiency virus type 1. *Mol Pharmacol*, **52**, 771-780.
38. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, De Clercq E, and Debyser Z (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem*, **278**, 372-381.
39. Chow SA (1997) In vitro assays for activities of retroviral integrase. *Methods*, **12**, 306-317.

40. Chow SA and Brown PO (1994) Substrate features important for recognition and catalysis by human immunodeficiency virus type 1 integrase identified by using novel DNA substrates. *J Virol*, **68**, 3896-3907.
41. Chow SA, Vincent KA, Ellison V, and Brown PO (1992) Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science*, **255**, 723-726.
42. Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey MA, Santos-Ferreira MO, Laurent AG, Dauguet C, Katlama C, Rouzioux C, and . (1986) Isolation of a new human retrovirus from West African patients with AIDS. *Science*, **233**, 343-346.
43. Cournaud V, Muller-Trutwin M, and Sonigo P (2004) [Evolution and virulence of primate lentiviruses]. *Med Sci (Paris)*, **20**, 448-452.
44. Cullen BR (1998) Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology*, **249**, 203-210.
45. Daniel R, Kao G, Taganov K, Greger JG, Favorova O, Merkel G, Yen TJ, Katz RA, and Skalka AM (2003) Evidence that the retroviral DNA integration process triggers an ATR-dependent DNA damage response. *Proc Natl Acad Sci U S A*, **100**, 4778-4783.
46. Daniel R, Katz RA, and Skalka AM (1999) A role for DNA-PK in retroviral DNA integration. *Science*, **284**, 644-647.
47. Davies DR, Goryshin IY, Reznikoff WS, and Rayment I (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science*, **289**, 77-85.
48. De Luca L, Pedretti A, Vistoli G, Barreca ML, Villa L, Monforte P, and Chimirri A (2003) Analysis of the full-length integrase-DNA complex by a modified approach for DNA docking. *Biochem Biophys Res Commun*, **310**, 1083-1088.
49. De Luca L, Vistoli G, Pedretti A, Barreca ML, and Chimirri A (2005) Molecular dynamics studies of the full-length integrase-DNA complex. *Biochem Biophys Res Commun*, **336**, 1010-1016.
50. Delelis O, Carayon K, Guiot E, Leh H, Tauc P, Brochon JC, Mouscadet JF, and Deprez E (2008) Insight into the integrase-DNA recognition mechanism: A specific DNA-binding mode revealed by an enzymatically labeled integrase. *J Biol Chem*.
51. Delelis O, Lehmann-Che J, and Saib A (2004) Foamy viruses--a world apart. *Curr Opin Microbiol*, **7**, 400-406.
52. Delelis O, Parissi V, Leh H, Mbemba G, Petit C, Sonigo P, Deprez E, and Mouscadet JF (2007) Efficient and specific internal cleavage of a retroviral palindromic DNA sequence by tetrameric HIV-1 integrase. *PLoS ONE*, **2**, e608.
53. Delelis O, Petit C, Leh H, Mbemba G, Mouscadet JF, and Sonigo P (2005) A novel function for spumaretrovirus integrase: an early requirement for integrase-mediated cleavage of 2 LTR circles. *Retrovirology*, **2**, 31.

54. Delelis O, Saib A, and Sonigo P (2003) Biphasic DNA synthesis in spumaviruses. *J Virol*, **77**, 8141-8146.
55. Deprez E, Barbe S, Kolaski M, Leh H, Zouhiri F, Auclair C, Brochon JC, Le Bret M, and Mouscadet JF (2004) Mechanism of HIV-1 integrase inhibition by styrylquinoline derivatives in vitro. *Mol Pharmacol*, **65**, 85-98.
56. Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, and Brochon JC (2000) Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. *Biochemistry*, **39**, 9275-9284.
57. Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, Hawkins ME, and Brochon JC (2001) DNA binding induces dissociation of the multimeric form of HIV-1 integrase: a time-resolved fluorescence anisotropy study. *Proc Natl Acad Sci U S A*, **98**, 10090-10095.
58. Devroe E, Engelman A, and Silver PA (2003) Intracellular transport of human immunodeficiency virus type 1 integrase. *J Cell Sci*, **116**, 4401-4408.
59. Dvorin JD, Bell P, Maul GG, Yamashita M, Emerman M, and Malim MH (2002) Reassessment of the roles of integrase and the central DNA flap in human immunodeficiency virus type 1 nuclear import. *J Virol*, **76**, 12087-12096.
60. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, and Davies DR (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981-1986.
61. Eijkelenboom AP, Lutzke RA, Boelens R, Plasterk RH, Kaptein R, and Hard K (1995) The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat Struct Biol*, **2**, 807-810.
62. Ellison V, Gerton J, Vincent KA, and Brown PO (1995) An essential interaction between distinct domains of HIV-1 integrase mediates assembly of the active multimer. *J Biol Chem*, **270**, 3320-3326.
63. Engelman A, Bushman FD, and Craigie R (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J*, **12**, 3269-3275.
64. Engelman A and Craigie R (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J Virol*, **66**, 6361-6369.
65. Engelman A and Craigie R (1995) Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity. *J Virol*, **69**, 5908-5911.
66. Engelman A, Englund G, Orenstein JM, Martin MA, and Craigie R (1995) Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J Virol*, **69**, 2729-2736.

67. Engelman A, Hickman AB, and Craigie R (1994) 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.
68. Engelman A, Mizuuchi K, and Craigie R (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*, **67**, 1211-1221.
69. Enssle J, Jordan I, Mauer B, and Rethwilm A (1996) Foamy virus reverse transcriptase is expressed independently from the Gag protein. *Proc Natl Acad Sci U S A*, **93**, 4137-4141.
70. Espeseth AS, Felock P, Wolfe A, Witmer M, Grobler J, Anthony N, Egbertson M, Melamed JY, Young S, Hamill T, Cole JL, and Hazuda DJ (2000) HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc Natl Acad Sci U S A*, **97**, 11244-11249.
71. Esposito D and Craigie R (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J*, **17**, 5832-5843.
72. Faure A, Calmels C, Desjobert C, Castroviejo M, Caumont-Sarcos A, Tarrago-Litvak L, Litvak S, and Parissi V (2005) HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res*, **33**, 977-986.
73. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R, Gallant J, Markowitz M, Ho DD, Richman DD, and Siliciano RF (1997) Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*, **278**, 1295-1300.
74. Freed EO (1998) HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology*, **251**, 1-15.
75. Gallay P, Hope T, Chin D, and Trono D (1997) HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc Natl Acad Sci U S A*, **94**, 9825-9830.
76. Gao K, Butler SL, and Bushman F (2001) Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes. *EMBO J*, **20**, 3565-3576.
77. Gao K, Wong S, and Bushman F (2004) Metal binding by the D,DX35E motif of human immunodeficiency virus type 1 integrase: selective rescue of Cys substitutions by Mn<sup>2+</sup> in vitro. *J Virol*, **78**, 6715-6722.
78. Geijtenbeek TB, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GC, Middel J, Cornelissen IL, Nottet HS, KewalRamani VN, Littman DR, Figdor CG, and van KY (2000) DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell*, **100**, 587-597.
79. Gelderblom HR, Hausmann EH, Ozel M, Pauli G, and Koch MA (1987) Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. *Virology*, **156**, 171-176.

80. Gerton JL, Herschlag D, and Brown PO (1999) Stereospecificity of reactions catalyzed by HIV-1 integrase. *J Biol Chem*, **274**, 33480-33487.
81. Goepfert PA, Wang G, and Mulligan MJ (1995) Identification of an ER retrieval signal in a retroviral glycoprotein. *Cell*, **82**, 543-544.
82. Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, and Davies DR (1998) Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc Natl Acad Sci U S A*, **95**, 9150-9154.
83. Graziosi C, Pantaleo G, Butini L, Demarest JF, Saag MS, Shaw GM, and Fauci AS (1993) Kinetics of human immunodeficiency virus type 1 (HIV-1) DNA and RNA synthesis during primary HIV-1 infection. *Proc Natl Acad Sci U S A*, **90**, 6405-6409.
84. Greene WC and Peterlin BM (2002) Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. *Nat Med*, **8**, 673-680.
85. Greenwald J, Le V, Butler SL, Bushman FD, and Choe S (1999) The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry*, **38**, 8892-8898.
86. Grobler JA, Stillmock K, Hu B, Witmer M, Felock P, Espeseth AS, Wolfe A, Egbertson M, Bourgeois M, Melamed J, Wai JS, Young S, Vacca J, and Hazuda DJ (2002) Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. *Proc Natl Acad Sci U S A*, **99**, 6661-6666.
87. Guiot E, Carayon K, Delelis O, Simon F, Tauc P, Zubin E, Gottikh M, Mouscadet JF, Brochon JC, and Deprez E (2006) Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. *J Biol Chem*, **281**, 22707-22719.
88. Ha HC, Juluri K, Zhou Y, Leung S, Hermankova M, and Snyder SH (2001) Poly(ADP-ribose) polymerase-1 is required for efficient HIV-1 integration. *Proc Natl Acad Sci U S A*, **98**, 3364-3368.
89. Harper AL, Skinner LM, Sudol M, and Katzman M (2001) 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.
90. Hayouka Z, Rosenbluh J, Levin A, Loya S, Lebendiker M, Veprintsev D, Kotler M, Hizi A, Loyter A, and Friedler A (2007) Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. *Proc Natl Acad Sci U S A*, **104**, 8316-8321.
91. Hayouka Z, Rosenbluh J, Levin A, Maes M, Loyter A, and Friedler A (2008) Peptides derived from HIV-1 Rev inhibit HIV-1 integrase in a shiftide mechanism. *Biopolymers*, **90**, 481-487.
92. Hazuda DJ, Anthony NJ, Gomez RP, Jolly SM, Wai JS, Zhuang L, Fisher TE, Embrey M, Guare JP, Jr., Egbertson MS, Vacca JP, Huff JR, Felock PJ, Witmer MV, Stillmock KA, Danovich R, Grobler J, Miller MD, Espeseth AS, Jin L, Chen IW, Lin JH, Kassahun K, Ellis JD, Wong BK, Xu W, Pearson PG, Schleif WA, Cortese R, Emimi E, Summa V, Holloway MK, and Young SD (2004) A naphthyridine

- carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc Natl Acad Sci U S A*, **101**, 11233-11238.
93. Heinzinger NK, Bukinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee MA, Gendelman HE, Ratner L, Stevenson M, and Emerman M (1994) The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc Natl Acad Sci U S A*, **91**, 7311-7315.
  94. Henriot S, Richer D, Bernacchi S, Decroly E, Vigne R, Ehresmann B, Ehresmann C, Paillart JC, and Marquet R (2005) Cooperative and specific binding of Vif to the 5' region of HIV-1 genomic RNA. *J Mol Biol*, **354**, 55-72.
  95. Heuer TS and Brown PO (1998) Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry*, **37**, 6667-6678.
  96. Hindmarsh P, Ridky T, Reeves R, Andrade M, Skalka AM, and Leis J (1999) HMG protein family members stimulate human immunodeficiency virus type 1 and avian sarcoma virus concerted DNA integration in vitro. *J Virol*, **73**, 2994-3003.
  97. Ho DD (1998) Toward HIV eradication or remission: the tasks ahead. *Science*, **280**, 1866-1867.
  98. Ho DD, Moudgil T, and Alam M (1989) Quantitation of human immunodeficiency virus type 1 in the blood of infected persons. *N Engl J Med*, **321**, 1621-1625.
  99. Hooks JJ and Gibbs CJ, Jr. (1975) The foamy viruses. *Bacteriol Rev*, **39**, 169-185.
  100. Huang M, Maynard A, Turpin JA, Graham L, Janini GM, Covell DG, and Rice WG (1998) Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. *J Med Chem*, **41**, 1371-1381.
  101. Jack WE, Terry BJ, and Modrich P (1982) Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence. *Proc Natl Acad Sci U S A*, **79**, 4010-4014.
  102. Jacque JM and Stevenson M (2006) The inner-nuclear-envelope protein emerlin regulates HIV-1 infectivity. *Nature*, **441**, 641-645.
  103. Jaroszewski JW, Syi JL, Maizel J, and Cohen JS (1992) Towards rational design of antisense DNA: molecular modelling of phosphorothioate DNA analogues. *Anticancer Drug Des*, **7**, 253-262.
  104. Jeltsch A, Wenz C, Stahl F, and Pingoud A (1996) Linear diffusion of the restriction endonuclease EcoRV on DNA is essential for the in vivo function of the enzyme. *EMBO J*, **15**, 5104-5111.
  105. Jenkins TM, Engelman A, Ghirlando R, and Craigie R (1996) 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.

106. Jenkins TM, Hickman AB, Dyda F, Ghirlando R, Davies DR, and Craigie R (1995) Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc Natl Acad Sci U S A*, **92**, 6057-6061.
107. Johnson AA, Santos W, Pais GC, Marchand C, Amin R, Burke TR, Jr., Verdine G, and Pommier Y (2006) Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J Biol Chem*, **281**, 461-467.
108. Jonsson CB, Donzella GA, and Roth MJ (1993) Characterization of the forward and reverse integration reactions of the Moloney murine leukemia virus integrase protein purified from *Escherichia coli*. *J Biol Chem*, **268**, 1462-1469.
109. Jordan A, Defechereux P, and Verdin E (2001) The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J*, **20**, 1726-1738.
110. Kalpana GV, Marmon S, Wang W, Crabtree GR, and Goff SP (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science*, **266**, 2002-2006.
111. Kammler S, Otte M, Hauber I, Kjems J, Hauber J, and Schaal H (2006) The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology*, **3**, 89.
112. Karki RG, Tang Y, Burke TR, Jr., and Nicklaus MC (2004) 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.
113. Kogel D, Aboud M, and Flugel RM (1995) Molecular biological characterization of the human foamy virus reverse transcriptase and ribonuclease H domains. *Virology*, **213**, 97-108.
114. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, and Hendrickson WA (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, **393**, 648-659.
115. Laboulais C, Deprez E, Leh H, Mouscadet JF, Brochon JC, and Le Bret M (2001) HIV-1 integrase catalytic core: molecular dynamics and simulated fluorescence decays. *Biophys J*, **81**, 473-489.
116. Lapadat-Tapolsky M, de Rocquigny H, Van Gent D, Roques B, Plasterk R, and Darlix JL (1993) Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res*, **21**, 831-839.
117. Lau A, Kanaar R, Jackson SP, and O'Connor MJ (2004) Suppression of retroviral infection by the RAD52 DNA repair protein. *EMBO J*, **23**, 3421-3429.
118. Lee HS, Kang SY, and Shin CG (2005a) Characterization of the functional domains of human foamy virus integrase using chimeric integrases. *Mol Cells*, **19**, 246-255.

119. Lee MC, Deng J, Briggs JM, and Duan Y (2005b) Large-scale conformational dynamics of the HIV-1 integrase core domain and its catalytic loop mutants. *Biophys J*, **88**, 3133-3146.
120. Lee SP, Xiao J, Knutson JR, Lewis MS, and Han MK (1997) Zn<sup>2+</sup> promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. *Biochemistry*, **36**, 173-180.
121. Leh H, Brodin P, Bischerour J, Deprez E, Tauc P, Brochon JC, LeCam E, Coulaud D, Auclair C, and Mouscadet JF (2000) Determinants of Mg<sup>2+</sup>-dependent activities of recombinant human immunodeficiency virus type 1 integrase. *Biochemistry*, **39**, 9285-9294.
122. Li M and Craigie R (2005) Processing of viral DNA ends channels the HIV-1 integration reaction to concerted integration. *J Biol Chem*, **280**, 29334-29339.
123. Li M, Mizuuchi M, Burke TR, Jr., and Craigie R (2006) Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J*, **25**, 1295-1304.
124. Limon A, Nakajima N, Lu R, Ghory HZ, and Engelman A (2002) Wild-type levels of nuclear localization and human immunodeficiency virus type 1 replication in the absence of the central DNA flap. *J Virol*, **76**, 12078-12086.
125. Lins RD, Adesokan A, Soares TA, and Briggs JM (2000) Investigations on human immunodeficiency virus type 1 integrase/DNA binding interactions via molecular dynamics and electrostatics calculations. *Pharmacol Ther*, **85**, 123-131.
126. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH, Teo W, and Poeschla EM (2006) An Essential Role for LEDGF/p75 in HIV Integration. *Science*.
127. Lochelt M and Flugel RM (1996) The human foamy virus pol gene is expressed as a Pro-Pol polyprotein and not as a Gag-Pol fusion protein. *J Virol*, **70**, 1033-1040.
128. Lochelt M, Muranyi W, and Flugel RM (1993) Human foamy virus genome possesses an internal, Bel-1-dependent and functional promoter. *Proc Natl Acad Sci U S A*, **90**, 7317-7321.
129. Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, and Gronenborn AM (1995) Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry*, **34**, 9826-9833.
130. Lubkowski J, Yang F, Alexandratos J, Wlodawer A, Zhao H, Burke TR, Jr., Neamati N, Pommier Y, Merkel G, and Skalka AM (1998) Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor. *Proc Natl Acad Sci U S A*, **95**, 4831-4836.
131. Lutzke RA and Plasterk RH (1998) Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. *J Virol*, **72**, 4841-4848.

132. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, Debyser Z, and Engelborghs Y (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem*, **278**, 33528-33539.
133. Maignan S, Guilloteau JP, Zhou-Liu Q, Clement-Mella C, and Mikol V (1998) Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J Mol Biol*, **282**, 359-368.
134. Malet I, Delelis O, Valantin MA, Montes B, Soulie C, Wirden M, Tchertanov L, Peytavin G, Reynes J, Mouscadet JF, Katlama C, Calvez V, and Marcelin AG (2008) Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. *Antimicrob Agents Chemother*, **52**, 1351-1358.
135. Malim MH, Hauber J, Le SY, Maizel JV, and Cullen BR (1989) The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature*, **338**, 254-257.
136. Marchand C, Johnson AA, Karki RG, Pais GC, Zhang X, Cowansage K, Patel TA, Nicklaus MC, Burke TR, Jr., and Pommier Y (2003) Metal-dependent inhibition of HIV-1 integrase by beta-diketo acids and resistance of the soluble double-mutant (F185K/C280S). *Mol Pharmacol*, **64**, 600-609.
137. Marchand C, Zhang X, Pais GC, Cowansage K, Neamati N, Burke TR, Jr., and Pommier Y (2002) Structural determinants for HIV-1 integrase inhibition by beta-diketo acids. *J Biol Chem*, **277**, 12596-12603.
138. Maroun RG, Gayet S, Benleulmi MS, Porumb H, Zargarian L, Merad H, Leh H, Mouscadet JF, Troalen F, and Femandjian S (2001) Peptide inhibitors of HIV-1 integrase dissociate the enzyme oligomers. *Biochemistry*, **40**, 13840-13848.
139. Mazumder A, Neamati N, Ojwang JO, Sunder S, Rando RF, and Pommier Y (1996) Inhibition of the human immunodeficiency virus type 1 integrase by guanosine quartet structures. *Biochemistry*, **35**, 13762-13771.
140. McDonald D, Vodicka MA, Lucero G, Svitkina TM, Borisy GG, Emerman M, and Hope TJ (2002) Visualization of the intracellular behavior of HIV in living cells. *J Cell Biol*, **159**, 441-452.
141. Miller MD, Farnet CM, and Bushman FD (1997) Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol*, **71**, 5382-5390.
142. Mirambeau G, Lyonnais S, Coulaud D, Hameau L, Lafosse S, Jeusset J, Borde I, Reboud-Ravaux M, Restle T, Gorelick RJ, and Le CE (2007) HIV-1 protease and reverse transcriptase control the architecture of their nucleocapsid partner. *PLoS ONE*, **2**, e669.
143. Moebes A, Enssle J, Bieniasz PD, Heinkelein M, Lindemann D, Bock M, McClure MO, and Rethwilm A (1997) Human foamy virus reverse transcription that occurs late in the viral replication cycle. *J Virol*, **71**, 7305-7311.

144. Molteni V, Rhodes D, Rubins K, Hansen M, Bushman FD, and Siegel JS (2000) A new class of HIV-1 integrase inhibitors: the 3,3,3', 3'-tetramethyl-1,1'-spirobi(indan)-5,5',6,6'-tetrol family. *J Med Chem*, **43**, 2031-2039.
145. Moreau K, Faure C, Violot S, Gouet P, Verdier G, and Ronfort C (2004) Mutational analyses of the core domain of Avian Leukemia and Sarcoma Viruses integrase: critical residues for concerted integration and multimerization. *Virology*, **318**, 566-581.
146. Moreau K, Faure C, Violot S, Verdier G, and Ronfort C (2003) Mutations in the C-terminal domain of ALSV (Avian Leukemia and Sarcoma Viruses) integrase alter the concerted DNA integration process in vitro. *Eur J Biochem*, **270**, 4426-4438.
147. Mousnier A, Leh H, Mouscadet JF, and Dargemont C (2004) Nuclear import of HIV-1 integrase is inhibited in vitro by styrylquinoline derivatives. *Mol Pharmacol*, **66**, 783-788.
148. Mulder LC, Chakrabarti LA, and Muesing MA (2002) Interaction of HIV-1 integrase with DNA repair protein hRad18. *J Biol Chem*, **277**, 27489-27493.
149. Mules EH, Uzun O, and Gabriel A (1998) In vivo Ty1 reverse transcription can generate replication intermediates with untidy ends. *J Virol*, **72**, 6490-6503.
150. Neamati N, Lin Z, Karki RG, Orr A, Cowansage K, Strumberg D, Pais GC, Voigt JH, Nicklaus MC, Winslow HE, Zhao H, Turpin JA, Yi J, Skalka AM, Burke TR, Jr., and Pommier Y (2002) Metal-dependent inhibition of HIV-1 integrase. *J Med Chem*, **45**, 5661-5670.
151. Nermut MV and Fassati A (2003) Structural analyses of purified human immunodeficiency virus type 1 intracellular reverse transcription complexes. *J Virol*, **77**, 8196-8206.
152. Pantaleo G, Graziosi C, Demarest JF, Butini L, Montroni M, Fox CH, Orenstein JM, Kotler DP, and Fauci AS (1993) HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature*, **362**, 355-358.
153. Perryman AL and McCammon JA (2002) AutoDocking dinucleotides to the HIV-1 integrase core domain: exploring possible binding sites for viral and genomic DNA. *J Med Chem*, **45**, 5624-5627.
154. Pettit SC, Lindquist JN, Kaplan AH, and Swanstrom R (2005) Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology*, **2**, 66.
155. Piatak M, Jr., Saag MS, Yang LC, Clark SJ, Kappes JC, Luk KC, Hahn BH, Shaw GM, and Lifson JD (1993) High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science*, **259**, 1749-1754.
156. Pinskaya M, Romanova E, Volkov E, Deprez E, Leh H, Brochon JC, Mouscadet JF, and Gottikh M (2004) HIV-1 integrase complexes with DNA dissociate in the presence of short oligonucleotides conjugated to acridine. *Biochemistry*, **43**, 8735-8743.

157. Podtelezhnikov AA, Gao K, Bushman FD, and McCammon JA (2003) Modeling HIV-1 integrase complexes based on their hydrodynamic properties. *Biopolymers*, **68**, 110-120.
158. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, and Gallo RC (1980) Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci U S A*, **77**, 7415-7419.
159. Pommier Y, Johnson AA, and Marchand C (2005) Integrase inhibitors to treat HIV/AIDS. *Nat Rev Drug Discov*, **4**, 236-248.
160. Ren G, Gao K, Bushman FD, and Yeager M (2007) Single-particle image reconstruction of a tetramer of HIV integrase bound to DNA. *J Mol Biol*, **366**, 286-294.
161. Roy J, Rudolph W, Juretzek T, Gartner K, Bock M, Herchenroder O, Lindemann D, Heinkelein M, and Rethwilm A (2003) Feline foamy virus genome and replication strategy. *J Virol*, **77**, 11324-11331.
162. Savarino A (2007) In-Silico docking of HIV-1 integrase inhibitors reveals a novel drug type acting on an enzyme/DNA reaction intermediate. *Retrovirology*, **4**, 21.
163. Schliephake AW and Rethwilm A (1994) Nuclear localization of foamy virus Gag precursor protein. *J Virol*, **68**, 4946-4954.
164. Semenova EA, Marchand C, and Pommier Y (2008) HIV-1 integrase inhibitors: update and perspectives. *Adv Pharmacol*, **56**, 199-228.
165. Sharp PM, Bailes E, Robertson DL, Gao F, and Hahn BH (1999) Origins and evolution of AIDS viruses. *Biol Bull*, **196**, 338-342.
166. Shehu-Xhilaga M, Hill M, Marshall JA, Kappes J, Crowe SM, and Mak J (2002) The conformation of the mature dimeric human immunodeficiency virus type 1 RNA genome requires packaging of pol protein. *J Virol*, **76**, 4331-4340.
167. Shimura K, Kodama E, Sakagami Y, Matsuzaki Y, Watanabe W, Yamataka K, Watanabe Y, Ohata Y, Doi S, Sato M, Kano M, Ikeda S, and Matsuoka M (2008) Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). *J Virol*, **82**, 764-774.
168. Shun MC, Daigle JE, Vandegraaff N, and Engelman A (2007a) Wild-type levels of human immunodeficiency virus type 1 infectivity in the absence of cellular emerlin protein. *J Virol*, **81**, 166-172.
169. Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, Kellam P, Cherepanov P, and Engelman A (2007b) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev*, **21**, 1767-1778.

170. Sinha S and Grandgenett DP (2005) Recombinant human immunodeficiency virus type 1 integrase exhibits a capacity for full-site integration in vitro that is comparable to that of purified preintegration complexes from virus-infected cells. *J Virol*, **79**, 8208-8216.
171. Sinha S, Pursley MH, and Grandgenett DP (2002) Efficient concerted integration by recombinant human immunodeficiency virus type 1 integrase without cellular or viral cofactors. *J Virol*, **76**, 3105-3113.
172. Smolov M, Gottikh M, Tashlitskii V, Korolev S, Demidyuk I, Brochon JC, Mouscadet JF, and Deprez E (2006) Kinetic study of the HIV-1 DNA 3'-end processing. *FEBS J*, **273**, 1137-1151.
173. Su PY and Brown TA (1997) Ty3/gypsy-like retrotransposon sequences in tomato. *Plasmid*, **38**, 148-157.
174. Surby MA and Reich NO (1996) Facilitated diffusion of the EcoRI DNA methyltransferase is described by a novel mechanism. *Biochemistry*, **35**, 2209-2217.
175. Suzuki Y and Craigie R (2007) The road to chromatin - nuclear entry of retroviruses. *Nat Rev Microbiol*, **5**, 187-196.
176. Topper M, Luo Y, Zhadina M, Mohammed K, Smith L, and Muesing MA (2007) Posttranslational acetylation of the human immunodeficiency virus type 1 integrase carboxyl-terminal domain is dispensable for viral replication. *J Virol*, **81**, 3012-3017.
177. Trono D (1992) Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses. *J Virol*, **66**, 4893-4900.
178. Turpin JA, Terpening SJ, Schaeffer CA, Yu G, Glover CJ, Felsted RL, Sausville EA, and Rice WG (1996) Inhibitors of human immunodeficiency virus type 1 zinc fingers prevent normal processing of gag precursors and result in the release of noninfectious virus particles. *J Virol*, **70**, 6180-6189.
179. van Gent DC, Oude Groeneger AA, and Plasterk RH (1993) Identification of amino acids in HIV-2 integrase involved in site-specific hydrolysis and alcoholysis of viral DNA termini. *Nucleic Acids Res*, **21**, 3373-3377.
180. Varmus H and Paul W (1997) AIDS: the process of discovery. *Science*, **275**, 141.
181. Vink C, Oude Groeneger AM, and Plasterk RH (1993) Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res*, **21**, 1419-1425.
182. Wang JY, Ling H, Yang W, and Craigie R (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J*, **20**, 7333-7343.
183. Wang LH, Yang XY, Zhang X, Mihalic K, Fan YX, Xiao W, Howard OM, Appella E, Maynard AT, and Farrar WL (2004) Suppression of breast cancer by chemical modulation of vulnerable zinc fingers in estrogen receptor. *Nat Med*, **10**, 40-47.

184. Weiss RA (2006) The discovery of endogenous retroviruses. *Retrovirology*, **3**, 67.
185. Wieggers K, Rutter G, Kottler H, Tessmer U, Hohenberg H, and Krausslich HG (1998) Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J Virol*, **72**, 2846-2854.
186. Wielens J, Crosby IT, and Chalmers DK (2005) A three-dimensional model of the human immunodeficiency virus type 1 integration complex. *J Comput Aided Mol Des*, **19**, 301-317.
187. Wiskerchen M and Muesing MA (1995) Identification and characterization of a temperature-sensitive mutant of human immunodeficiency virus type 1 by alanine scanning mutagenesis of the integrase gene. *J Virol*, **69**, 597-601.
188. Wu X, Li Y, Crise B, Burgess SM, and Munroe DJ (2005) Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J Virol*, **79**, 5211-5214.
189. Yang ZN, Mueser TC, Bushman FD, and Hyde CC (2000) Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J Mol Biol*, **296**, 535-548.
190. Yu SF, Baldwin DN, Gwynn SR, Yendapalli S, and Linial ML (1996) Human foamy virus replication: a pathway distinct from that of retroviruses and hepadnaviruses. *Science*, **271**, 1579-1582.
191. Yu SF, Sullivan MD, and Linial ML (1999) Evidence that the human foamy virus genome is DNA. *J Virol*, **73**, 1565-1572.
192. Zhang X, Pais GC, Svarovskaia ES, Marchand C, Johnson AA, Karki RG, Nicklaus MC, Pathak VK, Pommier Y, and Burke TR (2003) Azido-containing aryl beta-diketo acid HIV-1 integrase inhibitors. *Bioorg Med Chem Lett*, **13**, 1215-1219.
193. Zhao Z, McKee CJ, Kessl JJ, Santos WL, Daigle JE, Engelman A, Verdine G, and Kvaratskhelia M (2008) Subunit-specific Protein Footprinting Reveals Significant Structural Rearrangements and a Role for N-terminal Lys-14 of HIV-1 Integrase during Viral DNA Binding. *J Biol Chem*, **283**, 5632-5641.
194. Zheng R, Jenkins TM, and Craigie R (1996) 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.



## RESUME

L'intégration de l'ADN viral dans le génome des cellules hôtes est une étape obligatoire du cycle de réplication des rétrovirus. L'intégrase (IN) catalyse le processus global d'intégration en deux étapes distinctes et consécutives. La première des deux réactions, le 3'-processing, consiste en une coupure spécifique d'un dinucléotide au niveau des deux extrémités 3'-OH de l'ADN viral. L'IN transfère ensuite de manière concertée ces deux extrémités au sein de l'ADN cible. Nous avons, par des techniques basées sur l'anisotropie de fluorescence, caractérisé les paramètres structuraux et cinétiques du 3'-processing catalysé par l'IN du VIH-1. Nous avons montré d'une part que cette activité dépend de la taille des complexes IN / ADN. Nous avons trouvé que le dimère d'IN est la forme multimérique la plus active tandis que les complexes de haut poids moléculaire sont relativement peu efficaces. D'autre part, la structuration du domaine N-ter joue un rôle clé dans l'assemblage coopératif de ce dimère en présence de  $Mg^{2+}$  comme cofacteur cationique. Ce rôle n'est pas essentiel en présence de  $Mn^{2+}$ . L'étude de l'IN de PFV-1, un spumavirus, plus soluble, nous a permis de mettre en évidence une reconnaissance préférentielle de l'extrémité processée de l'ADN viral par cette IN, cette propriété n'avait jamais été observée pour l'IN du VIH-1 car elle était masquée par sa propension à l'agrégation. Nous avons également montré que ces deux IN rétrovirales réalisent le 3'-processing suivant un mécanisme catalytique lent de type « single turn-over ».

## ABSTRACT

The integration of the retroviral DNA into the host genome is a key process in the retroviruses life cycle. Integrase (IN) catalyzes two distinct and consecutive steps that constitute the overall integration process. The first of these two reactions is 3'-processing which corresponds to the specific cleavage of two nucleotides from each 3'OH-end of the linear viral DNA. Then, IN transfers both extremities into the target DNA, resulting in full-site integration. Using fluorescence anisotropy based assay, structural and kinetic parameters of the 3'-processing reaction catalysed by HIV-1 IN were characterized. We found that the activity was dependent on the molecular size of IN / DNA complexes. Large complexes were found to be relatively inefficient while the IN dimer was found to be the most active multimeric form. Moreover we found that the structural N-terminal domain rearrangement plays a key role for the cooperative DNA-binding mode of the IN dimer with  $Mg^{2+}$  as the cationic cofactor, but not with  $Mn^{2+}$ . In addition, a preferential interaction of IN with processed extremity of viral DNA interaction was clearly demonstrated for spumavirus PFV-1 IN, a more soluble IN; this property has never been observed for HIV-1 retroviral IN because of its aggregation propensity. For both retroviral IN, the 3'-processing reaction corresponds to a slow "single turn-over" mechanism.